



DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY CA 93943-5101

Approved for public release; distribution is unlimited.

**An Analysis of Traditional Operational Testing
to Project How Testing will be Conducted in the
Future**

by

**Lawrence L. Turner, Jr.
Captain, United States Army
B.S., Texas A&M University, 1983**

**Submitted in partial fulfillment
of the requirements for the degree of**

MASTER OF SCIENCE IN SYSTEMS ACQUISITION MANAGEMENT

from the

**NAVAL POSTGRADUATE SCHOOL
September 1993**

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE September 1993	3. REPORT TYPE AND DATES COVERED Master's Thesis, September 1991-September 1993
4. TITLE AND SUBTITLE An Analysis of Traditional Operational Testing to Project How Testing will be Conducted in the Future (U)		5. FUNDING NUMBERS
6. AUTHOR(S) Lawrence L. Turner, Jr.		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey CA 93943-5000		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSORING/MONITORING AGENCY REPORT NUMBER
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.		
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.		12b. DISTRIBUTION CODE A
13. ABSTRACT (maximum 200 words) The purpose of this thesis is to analyze current operational testing and to alert the reader to how testing will be conducted in the future. This analysis is conducted from a tester's point of view and is targeted at program offices so that Program Managers will be able to ensure that beneficial operational testing is conducted on their programs. The thesis also provides a detailed description of current and future systems used to instrument development systems, analyze the test data, and produce a final test report. It also provides a detailed analysis of a testing tool called Model-Test-Model (M-T-M). The goal of M-T-M is to save the Army time and money by using a validated model to provide operational test data. Finally, the thesis provides an analysis of the changes that will need to be made in the T&E community as a result of a new defense environment. These areas include the need for minimizing duplication between development and operational testing, using modeling and simulation, and taking advantage of other areas where testing costs can be reduced.		
14. SUBJECT TERMS Operational Testing, User testing, Model-Test-Model, Test Instrumentation, Real-Time Casualty Assessment, Future Testing		15. NUMBER OF PAGES 141
16. PRICE CODE		
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified
20. LIMITATION OF ABSTRACT UL		

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)

Prescribed by ANSI Std. Z39-18

ABSTRACT

The purpose of this thesis is to analyze current operational testing and to alert the reader to how testing will be conducted in the future. This analysis is conducted from a tester's point of view and is targeted at program offices so that Program Managers will be able to ensure that beneficial operational testing is conducted on their programs. The thesis also provides a detailed description of current and future systems used to instrument development systems, analyze the test data, and produce a final test report. It also provides a detailed analysis of a testing tool called Model-Test-Model (M-T-M). The goal of M-T-M is to save the Army time and money by using a validated model to provide operational test data. Finally, the thesis provides an analysis of the changes that will need to be made in the T&E community as a result of a new defense environment. These areas include the need for minimizing duplication between development and operational testing, using modeling and simulation, and taking advantage of other areas where testing costs can be reduced.

TABLE OF CONTENTS

I.	INTRODUCTION	1
A.	PURPOSE	1
B.	BACKGROUND	1
C.	THESIS OBJECTIVES	2
D.	RESEARCH QUESTIONS	2
E.	SCOPE	3
F.	METHODOLOGY	4
II.	OPERATIONAL TEST AND EVALUATION	6
A.	INTRODUCTION	6
B.	OPERATIONAL TEST AND EVALUATION	7
1.	Purpose	7
2.	OT&E Test Participants	7
3.	Program Management Office	8
4.	Test Coordinating Groups	9
5.	Types of Operational Test and Evaluation	10
6.	Early User Test and Evaluation (EUTE)	10
7.	Operational Assessments	11
8.	Initial Operational Test and Evaluation (IOT&E)	11
9.	Follow-On Operational Test and Evaluation	12
10.	Test Planning	12

11. Critical Operational Issues	13
12. Test Realism	13
13. Test Concept	14
14. Test Execution	14
15. Test Reports	15
C. PURPOSE OF TESTING	16
D. TEST AND EVALUATION INSTRUMENTATION	22
1. Position Location Systems	24
2. Simulated Fire System	27
3. Instrumentation Control	31
4. Support Instrumentation Systems	35
E. FUTURE INSTRUMENTATION	51
1. Mobile Army Field Instrumentation System (MAFIS)	52
2. Global Positioning System (GPS)	53
3. K-Band TRADOC Obscuration Pairing System (K-TOPS)	54
4. Non-Line-of-Sight Testing	54
F. Summary	55

III. PROBLEMS/CRITICAL QUESTIONS RAISED BY PROGRAM MANAGERS	57
A. INTRODUCTION	57
B. WHY TESTERS NEED ACCESS TO DEVELOPMENT SYSTEMS EARLY IN THE ACQUISITION CYCLE	57
1. SGT York	59

2. F-111, A-10, and A-4 Aircraft	60
3. Chaparral	60
C. WHY TEST TRIALS TAKE AS LONG AS THEY DO	61
D. THE FINAL TEST REPORT	62
 IV. MODEL-TEST-MODEL	65
A. MODELING INTRODUCTION/BACKGROUND	65
1. PURPOSES OF MODELING	66
2. TYPES OF MODELS	67
B. MODEL-TEST-MODEL	75
1. Phase 1: Long-Term Planning Phase	75
2. Phase II: Pretest Modeling	76
3. Phase III: Field Test Phase	78
4. Phase IV: Post-Test Modeling and Calibration Phase	79
5. Phase V: Model Validation	81
C. Model-Test-Model Using Janus	82
1. Janus Overview	82
2. Comparison of M1A1/M1A2 Battle Results Between Janus (A) and an Operational Field Test	83
3. Data Analysis	92
D. SUMMARY	98
 V. THE FUTURE OF TEST AND EVALUATION (T&E)	100
A. INTRODUCTION	100
B. CHANGES IN THE DEFENSE ENVIRONMENT	100

1. Weapons Requirements Process	103
2. Major Changes in the Research and Development (R&D) Process	104
3. Combined Industrial Base	105
4. Professional Acquisition Corps	106
C. ANALYSIS OF IMPACT OF CHANGES TO T&E	107
1. Information Technology	110
2. Simulation	111
3. Future Changes Require New Tools	113
4. Increased Diversity of Systems to be Tested	113
5. Impact on T&E Personnel	114
D. HOW THE TEST AND EVALUATION COMMUNITY SHOULD REACT TO CHANGES IN THE DEFENSE ENVIRONMENT .	114
1. Routine Functions	115
2. New Functions	117
E. SUMMARY	120
VI. CONCLUSIONS AND RECOMMENDATIONS	121
A. CONCLUSIONS	121
B. RECOMMENDED AREAS OF FUTURE RESEARCH	123
APPENDIX A. ABBREVIATIONS AND ACRONYMS	124
LIST OF REFERENCES	130
INITIAL DISTRIBUTION LIST	132

I. INTRODUCTION

A. PURPOSE

The purpose of this thesis is to analyze how traditional operational testing is conducted. The thesis also analyzes a tool called Model-Test-Model that is recently being used. Model-Test-Model provides a bridge between how testing is currently done and how testing will be conducted in the future. This analysis is conducted from a tester's point of view and is targeted at program offices so that Program Managers will be able to ensure that useful/beneficial operational testing is conducted on their programs.

B. BACKGROUND

Every procurement (major weapon system, communication systems, etc) has to demonstrate its benefits in terms of performance and cost. The best way of demonstrating the value and performance of a system is through testing. However, too many programs initially "spin their wheels" by not being sure how their particular system should be tested. Also, many program offices do not understand (and don't necessarily wish to understand) just what is involved in conducting a valid operational test. Additionally, with OSD (Office of the Secretary of Defense) mandating more simulation,

considerations need to be made as to what type and how much simulation to use. Are simulations/modeling really indicative of how a system will perform or how it might be produced? What are its benefits?

C. THESIS OBJECTIVES

The objective of this thesis is to provide future program managers insight and background information on operational testing and what the future role of operational testing will be. It also gives insights as to what tools may be available for the operational testing community.

D. RESEARCH QUESTIONS

The primary research question is:

"What do program offices need to know/consider in order to conduct operational testing of their programs?"

Subsidiary research questions are:

- What is operational testing and how is it done?
- Should program offices provide a representative system far in advance of the operational test?
- What is a "Countdown of Instrumentation"? Should program offices be concerned with it?
- What considerations of test equipment need to be made?
- What is Model-Test-Model and is it effective? Should program offices consider using it?
- What role does simulation play in operational testing?
- Should program offices be concerned with data reduction/analysis in advance of the operational test?

- What is the future of operational testing?

E. SCOPE

This thesis will analyze the conduct of operational testing using the Test and Experimentation Command (TEXCOM), Fort Hunter Liggett, California as a representative of Army test facilities. Fort Hunter Liggett (FHL) was chosen due to their key involvement in operational testing and because of their working relationship with the TRADOC Analysis Center (TRAC) - Monterey. Fort Hunter Liggett was also chosen as the model due to the researcher having intimate knowledge of its procedures and instrumentation. The researcher spent a tour at FHL as a Systems Engineer, Project Engineer, and as Chief of the Experimentation Engineering Branch. In these capacities, the researcher was required to interface with other test agencies such as White Sands Missile Range, Nellis Air Force Base (AFB) Test Agency, High Technology Test Bed at Fort Lewis, Operational Test and Evaluation Agency (OTEA), Harry Diamond Laboratories, Sandia Laboratories, Moffett Air Field, and the Test Agency at Fort Hood. The researcher was exposed to a variety of test facilities and FHL has comparable test instrumentation and facilities. Therefore, the researcher believes that FHL provides a valid representation of how operational testing is conducted and of what testing resources are available within the testing community.

Emphasis will be placed on providing agencies such as program offices an introduction into the "nuts and bolts" of how an operational test is conducted, why the tests are done a particular way, and what early considerations should be accommodated (first article weapon system, data reduction, etc.)

F. METHODOLOGY

Research for this thesis consisted primarily of an indepth literature review and interviews with key personnel involved with operational testing at Fort Hunter Liggett (FHL) and with modeling at the TRADOC Analysis Command (TRAC)-Monterey. The personnel interviewed at FHL included the Director of Operational Testing, the Chief of the Instrumentation Division, and the contractors involved with conducting the tests. These personnel were chosen in order to present the researcher with views of operational testing from top management, the people involved with putting instrumentation packages together, and the people responsible for actually conducting the tests. The Executive Officer at TRAC-Monterey was interviewed because he had experience at both conducting and managing the modeling efforts.

Government reports, instructions, directives, textbooks, and periodicals were used for information sources. A thorough review of operating procedures and on-site visits to Fort Hunter Liggett and TRAC-Monterey provided valuable information

on their current testing and modeling techniques, respectively. Additionally, the thesis researcher relied heavily on his past operational (Fort Hunter Liggett) and developmental (National Security Agency (NSA)) testing experiences.

II. OPERATIONAL TEST AND EVALUATION

A. INTRODUCTION

This chapter looks at the "nuts and bolts" of traditional operational testing and evaluation. It provides an introduction to the concept of Operational Test and Evaluation (OT&E). It gives the purpose of OT&E, presents the primary participants in the process, describes several types of OT&E and includes some guidelines for planning, execution, and reporting of OT&E programs. It is not meant as a rehash of previous publications that give an overall overview of how testing is conducted. Rather this chapter focuses on why the need for testing exists and on providing a detailed look at the types of testing resources/instrumentation that currently exist. This will provide program offices with invaluable information on what type of instrumentation exists to accurately represent and test their development system [Ref. 14]. This chapter also provides an analysis on the type of future instrumentation that will be part of the tester's inventory.

B. OPERATIONAL TEST AND EVALUATION

1. Purpose

Operational test and evaluation is conducted for major programs by an organization that is independent of the developing, procuring, and using commands. It is typically conducted in phases that are keyed to a decision review in the acquisition process. It is performed using user crews, operators or units in as realistic an environment as possible. The OT&E provides the decision authority with the estimate of [Ref. 3:p. 9-1]:

- The military utility, operational effectiveness and suitability of the new system.
- The system's desirability, considering systems already available, and the operational benefits or burdens associated with the new system.
- The need for modifications to the new system.
- The adequacy of doctrine, organizations, operating techniques, tactics, and training for employment of the system; the adequacy of maintenance support for the system; and the adequacy of the system's performance in the countermeasures environment.

2. OT&E Test Participants

In the Army, testing of major systems is accomplished by the Operational Test and Evaluation Command (OPTEC), which is an independent testing agency. Testing is conducted under conditions as operationally realistic as possible. Troops operating, maintaining, and supporting the system during testing are trained to the same level as troops who will actually perform these functions in the units. Program

management office personnel and test coordinating groups also play important roles in the overall OT&E process.

3. Program Management Office

The Program Manager (PM) plays an important role in the planning, reporting, and funding of operational testing. He/she must coordinate program activities with the test community. The PM also helps ensure that testing addresses the critical issues and provides feedback from testing activities to contractors.

At each milestone review, the PM is required to brief the decision authority on the testing planned and completed on the program. Therefore, it is important that the Program Management Office (PMO) personnel have a good understanding of the test program objectives. It is also important for them to work with the test community to ensure that the OT&E is well planned and adequate resources are available. The PMO involves the test community by organizing Test Coordinating Groups at the program initiation. It also involves the testing community by establishing channels of communication between the PMO and the key test organizations. The PMO staff should keep appropriate members of the test community well informed concerning system problems and the actions taken to correct them.

4. Test Coordinating Groups

The Army's Test Integration Working Group (TIWG), Navy's T&E Coordinating Group (T&ECG), and Air Force's Test Planning Working Group (TPWG) are chartered by their respective Service to coordinate and integrate the planning and execution of the T&E process [Ref. 3:p. 9-2]. The Army and Air Force groups are chaired by a representative of the PMO. The Navy's T&ECG is chaired by the development coordinator. The members of these groups represent the user, developmental tester, operational tester, independent evaluator, logistics, training, and contractors. The functions of the groups are to:

- Facilitate the use of testing expertise.
- Plan the use of instrumentation.
- Evaluate the types of Facilities needed.
- Determine which types of simulations and models to use.
- Integrate test requirements.
- Accelerate the Test and Evaluation Master Plan (TEMP) coordination process
- Resolve cost and scheduling problems
- Provide a forum to ensure that test and evaluation of the system is coordinated.

In the event of disagreement within a group, the issue is resolved through normal command/staff channels. In all of the Services, the groups help develop the TEMP.

5. Types of Operational Test and Evaluation

Operational Test and Evaluation can be subdivided into two phases: 1) operational testing performed before the full-rate production/deployment (Pre-production OT&E) decision and 2) the operational testing performed after the production decision [Ref. 3:p. 9-3]. The Pre-production OT&E includes Operational Assessments, Initial Operational Test and Evaluation (IOT&E) and Follow-On Operational Test and Evaluation (FOT&E). The operational assessments begin early in the program and continue until the system is certified ready for the independent operational test and evaluation. The IOT&E is conducted just prior to the full-scale production deployment decision. It continues until Initial Operating Capability (IOC) is achieved. After the full-rate production/deployment, all subsequent operational testing is referred to as Follow-On Operational Test and Evaluation.

6. Early User Test and Evaluation (EUTE)

Early User (Operational) Test and Evaluation is conducted primarily to forecast and evaluate the operational effectiveness and suitability of the weapon system during development. Operational assessments are conducted on the developing system until the PMO certifies that the prototype is ready for IOT&E.

7. Operational Assessments

Operational assessments begin after program initiation when the Operational Test Agencies (OTA) start their estimates of operational effectiveness and suitability. The OTA uses any testing results and data from other sources during an evaluation. As the program matures, these operational assessments are conducted on prototypes and preproduction articles. The assessments continue until the system is fully developed and certified ready for its IOT&E.

8. Initial Operational Test and Evaluation (IOT&E)

The IOT&E is the final dedicated phase of OT&E preceding a full-rate production decision. The IOT&E is conducted by an operational test and evaluation agency independent of the contractor, PMO, or Developing Agency. DOD Directive 5000.3 defines the test conditions under which such testing must be conducted:

Operational testing shall be accomplished in an environment as operationally realistic as possible, including threat representative hostile forces. Typical users should operate and maintain the system under conditions simulating combat stress and peacetime conditions.

The IOT&E must be conducted without system contractor personnel participation as set forth in Public Law 99-661 by Congress. The results from this test are evaluated and presented to the decision authority prior to the decision to enter full-rate production. This phase of OT&E addresses the

critical issues identified in the Decision Coordinating Paper (DCP) and the TEMP.

9. Follow-On Operational Test and Evaluation

The FOT&E is conducted after the Milestone III decision. Typically FOT&E is conducted using production systems. Specific objectives of FOT&E include:

- Testing of modifications that are to be incorporated into production systems.
- The completion of any deferred or incomplete IOT&E.
- Assessment of operational availability.
- Spares support.

10. Test Planning

Test planning is probably one of the most, if not the most, important parts of the OT&E process. Deliberate and complete planning may not guarantee a successful test program, but inadequate planning will result in significant test problems, system failure, and cost overruns. Operation test planning is performed by the Operational Test Agencies after program initiation prior to each operational test phase.

Operational test planning is divided into three phases: 1) Early planning, 2) advanced planning, and 3) detailed planning. Early planning involves critical operational issue development, determining the concept of operation, envisioning the operational environment, and developing mission scenarios and resource requirements. Advanced planning entails the determination of the purpose and

scope of testing, identification of critical issues, development of test objectives, establishment of a test approach and estimating test resource requirements. Detailed planning involves the development of step-by-step procedures to be followed as well as the coordination of resource requirements necessary to carry out OT&E.

11. Critical Operational Issues

One of the primary purposes of OT&E is to resolve critical operational issues about the system. The first step in an OT&E program is to identify these critical issues. Critical issues provide focus and direction for the operational test. When critical issues are properly addressed, deficiencies in the system can be uncovered and corrected. The issues form the basis for a structured technique of analysis by which detailed subobjectives (Measures of Effectiveness (MOE)) can be established. During an operational test, each subobjective is addressed by an actual test measurement. After these issues are identified, the evaluation plans and test design are developed for test execution.

12. Test Realism

Realism in an OT&E program includes all of the characteristics that make the test simulate actual combat conditions. There must be a concern for realism throughout the planning and conduct of the test. The three basic areas

of particular significance in applying realism identified by Roger Smith in his book, "Operational Test and Evaluation: A Systems Engineering Approach", are:

- During development of the test concept paper and design of the overall aspects of the test program, the developers must ensure the basic test philosophy is determined and realism is closely woven into this design.
- During planning and design of the actual test and development of scenarios, the planners must ensure that realism is included into the operational and maintenance activities.
- During the actual conduct of the tests, the field testers must ensure that the tactical realism is maintained.

13. Test Concept

In developing a test concept, it must be determined if OT&E will be performed in parallel with systems development, if all testing is to be done on production equipment, if testing will be evolutionary, and if testing will have to wait until all system capabilities are developed. These determinations can best be answered by considering a number of systems aspects such as test information requirements, system availability during test periods, and the implementation of system capabilities. The test concept is driven by the acquisition strategy and is used for test design and evaluation.

14. Test Execution

Test execution is the essential bridge between test planning and test reporting. For successful execution of the OT&E plan, the test officer must direct and control the test

resources and collect the data required for the evaluator to present to the decision authority. The test officer must prepare for testing, activate and train the test team, develop test procedures and operating instructions, control data management, create OT&E plan revisions, and manage each of the test missions. His/her data management duties will include raw data collection, creating a data status matrix, ensuring data quality assurance, processing and reduction, verification, filing, storage, retrieval, and analysis. Upon conclusion of all the test trials and the data reduction and analysis, the test results must be reported.

15. Test Reports

The OT&E test reports written for decision authorities must be timely, factual, concise, comprehensive and accurate. The report must present a balanced view of the weapon system's successes and failures during testing. It must also present the system's positive aspects and its deficiencies.

The four types of reports most frequently used in reporting OT&E results include: 1) status, 2) interim, 3) quick-look, and 4) final reports. The status report gives periodic updates (e.g., monthly, quarterly) and reports recent test findings. The interim report provides a summary of the cumulative test results to date. The quick-look report provides preliminary test results, is usually prepared immediately after a test event (less than 7 days) and may be

used to support program decision milestones due to the need to support a decision before the final report can be written. The final test report presents the final test results, conclusions, and recommendations covering the entire OT&E program with all supporting data.

C. PURPOSE OF TESTING

The materiel acquisition process can take many years from the time a materiel requirement is identified until the system is fielded. Although in this process operational (user) testing accounts for only a short time period, the results weigh heavily on any decisions to continue development, accept the system, or terminate the program for a system acquisition or to change organization, doctrine, and concepts for nonmateriel requirements [Ref. 4:p. 1-1]. Therefore, the fundamental purpose of test and evaluation (T&E) in a defense system's development and acquisition program is to identify the areas of risk to be reduced or eliminated [Ref. 3:p. 1-1].

Testing is an information generation activity with the objective of reducing the risk of doing something [Ref 6:p. 1]. In general terms, we test to generate information to reduce the risk in applying new technology or in using old technology in new ways. We stop testing when that risk has reduced to a level generally acceptable to those responsible for the application. Since the uncertainties introduced by the new technology or its novel applications drive test and

evaluation, testers tend to be challenged most by the very features and characteristics that make new systems effective, low observables technology being perhaps the most obvious example [Ref 6:p. 1]. The Defense Science Board recently published lists of High Leverage Technologies, Core Technologies, and Emerging Technologies, which, when cross-referenced against the list of Major Long-Term Military Goals, yields an intimidating matrix of test and evaluation challenges.

Some areas that are currently challenging the T&E community include the following:

- Nondevelopmental Items (NDI).
- Command and Control/Data Fusion.
- Strategic and Space Defense.
- Use of Computer Simulations.
- Threat Identification/Interpretation.
- Communication with Oversight Clientele.
- Software.
- Mission-Level Measures of Effectiveness.
- Modification/Regression Testing.
- Low Observables/Counter Low Observables.
- Directed Energy.
- Antisubmarine Warfare.
- Integrated Avionics.
- Survivability.
- Electronic Combat.

- Artificial Intelligence.
- Chemical/Biological/Radiological Weapons.
- Assessment of Real-World Data.
- Cost versus Sufficiency Testing.

In some of these areas, technical considerations are clouded by political, social or cultural ones. Both testers and developers have to be sensitive about how weapons technology is viewed. If, for example, we develop directed energy weaponry to defeat optics on the battlefield, how will the "public" react. There would exist the potential of having enemy combatants "blinded" by the use of this weaponry. This would go against our cultural psyche of having "clean" battles and the thought of us blinding soldiers on the battlefield would elicit strong objections from several groups. The technology to simulate directed energy weapons is complex and, when you add all the classified security measures that would have to be used to avert negative public opinion, you can see how this would severely hamper the testing effort. Another example concerns the use of lasers. I was a Project Engineer at FHL during a particular test. We were using our normal "eye-safe" lasers to simulate the firing of the system being tested. Somehow rumors were being circulated in the surrounding communities that we were using high-powered lasers and that the animals indigenous to the area were being blinded by the reflected laser energy. Hysteria then began to run rampant throughout the surrounding communities that reflected

laser energy would start blinding the people. We were ordered to halt testing until an investigation could be conducted. The investigation confirmed that we were not using high-powered lasers and we were allowed to continue our testing. However, the trials we were doing at that point were invalid and had to be performed again. This was in addition to the days we had already lost due to the investigation. Therefore, the testing schedule had to be prolonged, which added significant cost to the test and delayed the writing of the final test report. This delay slowed down the acquisition cycle. It also affected the morale of the user troops who were on temporary duty (TDY) and had already spent a considerable time away from their home and families.

It has been said that, despite its utility, "Testing is a zero value-added activity", since no system is ever improved by the act of testing alone [Ref. 6:p. 2]. Testing is usually performed either to find out something (experiment/test) or to prove something (demonstration) of either an operational or technical nature. The ultimate goal of the testing is to reduce the risk of an unwanted result of taking some action. Although, ideally, testing proceeds until adequate risk reduction has been achieved, numerous other constraints tend to affect the type, amount, and duration of testing:

- Resource Constraints - What type of testing range do we need and is it available? Does the technology exist to test the system we have? Is our testing budget sufficient to adequately test our system?

- Schedule Constraints - Is the test range that we need available and for how long? What season(s) of the year do we need to test?
- Environmental and Safety Constraints - Do we need to "live fire" test our system? Are we going to use laser designators or laser-guided weapons?
- Pressing need to Use, Deploy, or Market - How soon does this system need to go into production? Will this system be used in a current regional conflict?
- Desire to exploit a Technique, Technology or Opportunity- What are the minimum testing requirements? Do we need to operationally test or is modeling and simulation adequate?
- Security limits on Capability Exposure - What type of classification security requirements need to be met? Do we have adequate security safeguards to protect the data? How do we extensively coordinate with other agencies and maintain security?
- Programmatic Perturbations - How many systems on the testing master schedule will have their funding terminated? How much and what type of testing should be done on a system that may be terminated?.
- Political, Social, and Cultural Considerations - What type of administration is currently in office? How is public opinion in regards to the military environment? Are we likely to have protesters marching outside our testing facilities due to the testing of a certain weapon system?

In addition, what constitutes enough testing is dependent on the stakeholder(s). A program of any size will normally have a diverse set of stakeholders, each with a view on the sufficiency of testing. Based on my observations of the testing process, I conclude that the view of each stakeholder is as follows:

- The Developing Agency - Typically believe that they have sufficiently tested the system and that little to no further testing is needed.

- The Program Manager - Typically views the testing community as an agency that is looking for faults in his/her system.
- The Contractor - Their incentive is to go into production to generate a profit from their system. Therefore, regard testing as a necessary evil.
- The User - Typically wants the system "yesterday."
- The Tester - Typically have viewed their role as the "final exam" for the system. The system must "graduate" from their testing to go into production. Guardians of the public trust.
- The Review and Audit Community - Was everything done according to the book?
- The Analytical Community - Typically want to make sure they have enough data to make a correct analysis. However, how much testing is enough?
- The Taxpaying Public - The best system for the least amount of money.
- The Media - Guardians of the taxpayer. Bad news makes good news.
- Congress - Proponents for and against the system. Proponents for the system argue for limited testing while proponents against the system argue for extensive testing.

That each stakeholder has a right to his or her special perspective is not disputed. However, the tester is expected to supply special contributions, including independence and technical expertise. In other words, the tester is expected to maintain his/her independence while taking into account all of the stakeholders' concerns and magically producing the test plan that will have just the right amount of testing.

Discussions of how much testing is enough have traditionally focused on analytical approaches to the determination of test sample size. Statistical considerations

are essential, but may not embody all of the real-world factors. It is difficult to find literature identifying programs in which the amount of testing was precisely enough. Instances of insufficient testing are much easier to cite. Regardless, part of the professional tester's job is to determine how much testing is enough, assess and articulate the adequacy of testing, and defend his vision of testing sufficiency against those of other constituencies. This is to be done, furthermore, under very general test and evaluation policy guidance, directives, and regulations that are virtually free of any type of counsel regarding how much is enough.

D. TEST AND EVALUATION INSTRUMENTATION

This section provides basic descriptions, capabilities, and limitations of the major systems and equipment used by the TEXCOM Experimentation Center (TEC), formerly the Combat Development Experimentation Command (CDEC), at Fort Hunter Liggett, CA. The instrumentation used by TEC provides a good example of the resources available at major Army test facilities. By knowing what type of instrumentation is available and their capabilities and limitations, program offices should be better able to work with the operational testers in ensuring that their systems are accurately portrayed.

The TEC is capable of providing instrumentation for two types of tests [Ref. 11:p 1-1]:

- **One-Sided Live and Non-Live Fire Tests** - Conducted to provide a limited specific piece of data needed in a study or computer model. There is no standard test package for these tests and the instrumentation varies from simple to complex systems designed specifically for a particular test.
- **Force-on-Force Real-Time Casualty Assessment Tests** - Highly realistic mock battles in which casualties are assessed by a computer acting in the capacity of a high-speed umpire. These tests can be conducted during the day or night with individual weapons, combat vehicles, fixed-wing aircraft, helicopters, crew-served weapons or infantry indirect fire.

The instrumentation used to perform the two types of tests at FHL includes everything from a stopwatch to a complex computer system. This instrumentation is divided into four general categories:

- **Position Location Systems** - The Range Measuring System (RMS) acquires the ranges of soldiers, ground vehicles, and aircraft in relation to the known interrogator locations. The ranges are used by the computer to determine the position of the test players as a function of time.
- **Simulated Fire System** - Simulates firing at live targets and allows real-time scoring through the use of Direct Fire Simulators (DFS).
- **Instrumentation Control** - Consists of control equipment that causes subsystems of the instrumentation packages to react. The Instrumentation Computer Network (ICN) and the Integrated Information Control Center (I^2C^2) serve as the central control and data collection points for other instrumentation systems.
- **Support Instrumentation Systems** - Provide general support and utility in meeting data collection requirements. These systems include the Range Timing System (RTS), the Voice Recording System (VRS), the Range Communications system (RCS), and the Video System.

Figure 1 provides an overview of the TEC instrumentation. Each of the equipments that make-up the instrumentation categories will be discussed in further detail.

1. Position Location Systems

a. Range Measuring System

The RMS provides a position location capability when interfaced with the Computer Data Link (CDL) with the ICN. The RMS also provides a two-way telemetry link for data transmission and collection to and from a field player. The system was conceived as a means of rapidly determining the

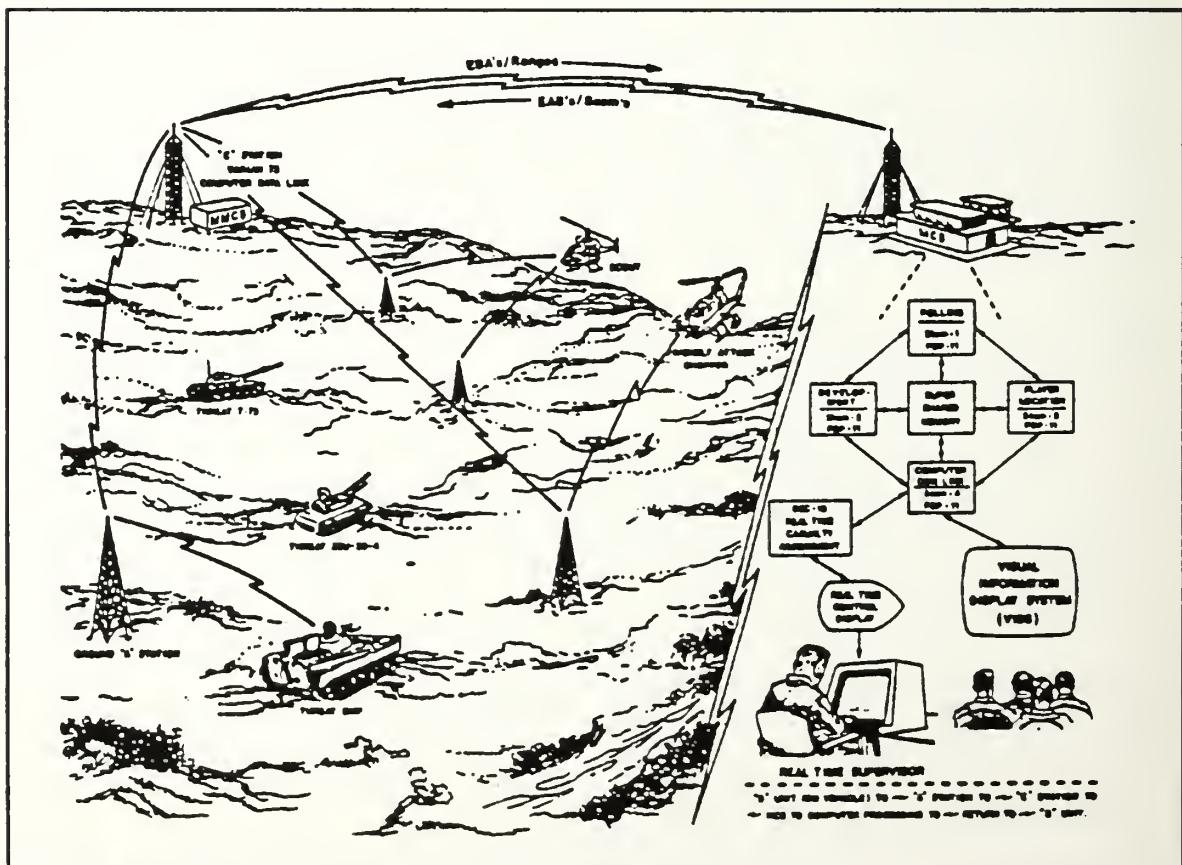


Figure 1 RMS and ICN at TEC

time-correlated position of soldiers, vehicles, and aircraft relative to each other in mock battles.

(1) Capabilities

- Manual Operation - The manual Control Panel forms a message containing the address of the unit desired and an operational mode bit. A range limit bit is then transmitted followed by a range pulse in range or data in SCOM (Short Communication Messages) and EAB (Extended A-Station to Micro-B unit).
- Computer Control Operation - This is identical to manual control except that the computer forms the message. In the event that the RF (Radio Frequency) link is broken, a no-response signal is generated after a time delay since the last micro-b unit responded.
- Ranging Mode - When a ranging message is sent to an A-Station, the A-Station sends a range pulse to the B-Unit, which waits 109 seconds and returns a range pulse to the A-Station. The 109 seconds is the B-Unit processing and transmission time. The A-Station compensates for the delay and transmits a number (range in two meter increments) to the computer via the C-Station. After the A-Stations (minimum three and maximum eight) range on a given B-Unit, the computer calculates the position of the B-Unit (player).
- Short Communication Messages (SCOM) - This mode can be used to communicate data to and from a B-Unit. Four data bits are transmitted from the C-Station (master transmitting station) to A-Station to B-Unit. The I/O (Input/Output) device can then transmit 13 data bits through the B-Unit to the A-Station to the C-Station to the computer.
- Extended A- to B- Mode (EAB) - This mode allows the computer to send a 42-bit message to the I/O device connected to the B-Unit. The B-Unit then acknowledges receipt of the message via the A- and C-Stations.
- Extended B- to A- Mode (EBA) - The computer can request a message from the I/O device, and the B-Unit will respond with a 42-bit message via the A- and C-Stations.

(2) Limitations

- The RF frequency for all RMS transmissions is 918 MHz. At this frequency, all communications require line-of-sight, which is difficult in the hilly terrain at FHL. The D-Station is used to overcome line-of-sight problems between the C-Station and A-Stations.
- The use of a single frequency band limits the system to only one transmission at a time, thereby limiting the number of players that can be polled in a given period of time. The theoretical upper limit is 1,686 messages per second. The practical upper limit is on the order of 1,000 messages per second.

b. Vestpack System

The individual-carried vestpack instrumentation system consists of an Upgraded Logic Module (ULM), a Speech Synthesizer/Limited Expansion Interface (SS/LEI) box, a Micro-B, and batteries. To complete the system, body-group sensors and a miniature speaker are attached to the vestpack, and head-group sensors and antennas are mounted on the player's helmet.

(1) Capabilities

The vestpack is capable of decoding both Multiple Integrated Laser Engagement System (MILES) and TEXCOM FHL laser codes.

(2) Limitations

The vestpack battery life with two 12-volt nickel cadmium battery packs is 13.5 hours.

c. Linear Triaxial Accelerometer (LTA)

The LTA is used as part of an instrumentation system that collects attitude and motion data from rotary- and fixed-wing aircraft. The LTA provides outputs representing the aircraft acceleration along three orthogonal axes aligned with the airframe.

(1) Capabilities

The LTA will measure acceleration along the three orthogonal axes to a maximum range of +/- five gravity.

(2) Limitations

Accelerometers will measure only up to +/- five g. This is sufficient for helicopters but would not be sufficient range for high-performance aircraft.

2. Simulated Fire System

a. Laser Transmitters

All Direct Fire Simulator (DFS) system laser transmitters employ Gallium Arsenide (GaAs) laser diodes as the source of the laser beam. The GaAs lasers emit invisible, near-infrared radiation of 904 nanometers (nm) wavelength. All transmitters have an optical system to produce nearly collimated laser beam output. Beam divergence angles range from 2.4 milliradians (mrad) to 52 mrad, depending on the laser transmitter type and application.

(1) Capabilities

- Upgraded Transmitter Signal (UTS) Laser Transmitter - the UTS laser transmitter can be used to simulate fire from infantry rifles, rifle grenades, machine guns, short range missiles, and large direct fire guns. Maximum laser-to-sensor pairing range is approximately 3,000 meters (m).
- IDFSS Laser Transmitter - Can be used to simulate fire from infantry rifles, rifle grenades, machine guns, short-range missiles, and large direct fire guns. Maximum laser-to-sensor pairing range is approximately 3,500 m.
- Schwartz Electro-Optical Small Gun Laser (SEO SGL) Transmitter - Can be used to simulate fire from the same weapon systems as the IDFSS laser transmitter. Its range is also 3,500 m.
- Advanced Anti-Armor Vehicle Evaluation (ARMVAL) Direct Fire Simulator (ADFS) Laser Transmitter - Can simulate fire from large weapons or missiles and can be vehicle-, aircraft-, or ground-mounted. Its range is in excess of 5,000 m.
- SEO Large Gun Laser (LGL) Laser Transmitter - Can be used to simulate fire from the same weapon systems as the ADFS laser transmitter. Its range is also in excess of 5,000 m.
- Air Defense Weapon Fire Simulator (AWFS) Laser Transmitter- Can simulate ground-to-air missile fire from air defense systems. It can also be used for ground-to-ground large weapons fire. The AWFS can simulate the target lockon angles of missile systems by scanning through sectors of three degrees in elevation by three degrees in azimuth. Its range is in excess of 5,000 m and can approach 10,000 m.

(2) Limitations

- The effective range of DFS laser beams can be impaired by smoke or dust clouds between laser and target.
- UTS Laser Transmitter - Potentially hazardous levels of optical radiation exist out to a range of five meters to the unaided eye, and 35 m if magnifying optics are used.
- IDFSS Laser Transmitter - Eye safety hazards can be encountered if personnel stare into the laser beam for

periods of time or through stabilized binoculars at ranges less than 10 m.

- SEO SGL Laser Transmitter - Potentially hazardous levels of optical radiation exist out to a range of five meters to the unaided eye, and 35 m if magnifying optics are used.
- ADFS Laser Transmitter - Potentially hazardous levels of optical radiation exist out to a range of 12 m to the unaided eye, and 260 m if magnifying optics are used.
- SEO LGL Laser Transmitter - Potentially hazardous levels of optical radiation exist out to a range of 15 m to the unaided eye, and 100 m if magnifying optics are used.
- AWFS Laser Transmitter - Potentially hazardous levels of optical radiation exist out to a range of 56 m to the unaided eye, and 750 m if magnifying optics are used.
- Laser Transmitter with Microcomputer (MCU) Backpack - When used by an infantry player, there is a limit to the time the laser and the rest of the MCU backpack can be continuously powered.

b. Laser Sensors

The laser sensor is the device that detects a hit from a laser transmitter. It consists of a near-infrared-sensitive photodiode and amplifiers packaged in a metal case. In front of the photosensitive area of the photodiode is a glass window, a mesh screen for reduction of electromagnetic interference, and an optical bandpass filter for improved signal-to-noise characteristics. The laser sensors can be mounted on infantry players, ground vehicles, and aircraft.

(1) Capabilities

- IDFSS Sensors - Designed for infantry players and can be used in some vehicle applications.
- Modified TEC Sensors - The modified sensors are used in vehicle applications. They have dual connectors so that

multiple sensors can be strung together on a single cable run. They can be mounted in a small cluster high on the vehicle, or distributed at separate locations lower on the vehicle. Helicopters can be instrumented with a small sensor cluster on each side.

- Simulaser SR104 Sensors - These sensors are smaller and lighter than the previous two sensors, but still have the same capabilities. Their small size and ease of mounting make these sensors ideal for incorporation into specialized detector assembling such as limited field-of-view detectors.

(2) Limitations

- Sensor windows must be kept clean of dust or mud accumulation.
- On vehicles, the sensors must be mounted free from field-of-view obscuration by vehicle appendages.

c. Gun Azimuth System (GAS)

The GAS senses the angle of the M-60 tank turret relative to the hull of the tank.

(1) Capabilities

The turret angle can be measured to an accuracy of ± 0.2 degrees. The angle can be transmitted to the ICN up to four times per second.

(2) Limitations - None.

d. Boresight Devices

Boresight devices are used to boresight the Direct Fire Simulator (DFS) laser beam to the sight of the weapon on which the laser is mounted. When hit by a laser beam, a laser sensor on the boresight device activates a light that signals the hit to the weapon operator. Two types of boresight

devices are used: the International Laser System (ILS) boresight target and the boresight box.

(1) Capabilities

- ILS Boresight Target - The ILS boresight target is used to align the laser on infantry rifles. Elevation and azimuth sighting errors are shown by an array on 12 sensors and lights on a circular target that can be up to 44 inches in diameter. The laser can be aligned to the rifles by adjusting the rifle sight until the laser is zeroed-in on the target center.
- Boresight Box - The boresight box is used to align the laser to the larger weapons. The boresight box contains a single sensor, electronic circuitry, and a single elevation/azimuth, adjustable automotive headlight. The sensor is on a long cable and can be temporarily affixed to the center of a large sighting target. The laser beam can be centered on the sensor by a scanning procedure, and the weapon's optical sight can then be adjusted to the center of the sighting target.

(2) Limitations

For optimum visibility, the headlight in the boresight box must be accurately aimed at the weapon operator.

3. Instrumentation Control

a. Instrumentation Computer Network (ICN)

The ICN's computers provide real-time control, data collection, and processing. In addition to controlling the Range Measuring System (RMS), the ICN processes data in real-time the position location (PL) of all players and assesses the engagements of field test players of opposing forces. It also supports the time-sharing for software development, data

reduction, and data reduction. The ICN is composed of four subsystems:

- The Digital Equipment Corporation (DEC) 1095 (DEC-10KL) Computer.
- The PDP-11/45 Computers or Mini-Computer Complex (MCC).
- The Mobile Multi-Purpose Control Station (MMCS).
- The MODCOMP II Control Computer of the Visual Information Display System (VIDS).

(1) Capabilities

- ICN support for large-scale Real-Time Casualty Assessment (RTCA) field trials includes weapon system engagement assessment and detailed data displays that are used to control the quality of field test data.
- ICN supports field tests involving up to 100 players (B-Units - small transponders that give each player a unique identity), depending on the types of players. The system also uses 84 RMS A-Stations (antenna towers used to relay telemetry data between the players and the computer). Additionally, a polling routine selects A-Stations on the basis of response, optimizing the use of A- to B- links.
- In special cases, Position Location (PL) data are within one to two meters. Typical PL data are accurate within five to 10 meters. PL accuracy depends on trade-offs between the number of players, the types of players and the polling rate.
- The system provides status information about each player (e.g., switches, lights, power, laser, and continuity) using firmware routines built into the player hardware. This information significantly reduces the time required to maintain and "countdown" before each trial (more will be said on just what a countdown is and what is involved).
- The Real Time Casualty Assessment (RTCA) subsystem consists of the following programs: 1.) **Direct Fire Casualty Assessment** (models main-gun weapons and short-range fire-and-forget weapons) 2.) **Missile Fire Casualty Assessment** (models command-guided and wire-guided missiles) 3.) **Automatic Gun Casualty Assessment** (models rapid fire weapons) 4.) **Small Arms Casualty Assessment** (models small arms) 5.) **Indirect Fire Casualty Assessment**

(models artillery weapons, including the Copperhead) 6). **ZSU Casualty Assessment** (models air defense guns including the ZSU) 7). **Hellfire Casualty Assessment** (models the Hellfire and Copperhead missiles)

(2) Limitations

- The ICN maximum player load depends on the type of players, data to be collected, and accuracy required. Maximum player load is also limited by the RMS's capacity of 1,000 messages per second. A typical player load is 40 to 60 players.
- Simultaneous experiment support is limited and requires careful coordination. The system cannot handle real-time operations and instrumentation checkout at the same time. The two experiments must share instrumentation checkout time and real-time operations.

b. Integrated Information Control Center (I²C²)

The I²C² configuration includes a Real-Time Computer Controller (RTCC), threat (red) and friendly (blue) controls, red and blue Indirect Fire Casualty Assessment System (IFCAS) controls, and communications and test director control areas. It also includes the VIDS theater. Although the I²C² has several functions, depending on the test being conducted, its primary functions are to provide RTCC and to exercise control and monitoring.

(1) Capabilities

- Monitoring of player movements.
- Display of firing engagements.
- Display of player statistics.
- Display of range information.
- Data recording.

- Variable-speed playback for post-trial analysis.

(2) Limitations

The lack of available space in the triple-wide trailer makes it increasingly difficult to expand the I²C² capabilities. There are only 12 pairs of telephone lines to control remote radios at Sites 8A and 8X. Considerable crosstalk on these telephone lines causes interference with communications.

c. Computer Data Link (CDL)

The CDL enables the Mobile Multi-Purpose Control Station computers, located at site 8A (on top of a hill), to communicate with the ICN (located five kilometers away in the base computer complex). The CDL consists of two Collins Mar-1108 FM (frequency modulated) microwave transmitter/receiver systems. One system is located at each end of the link. The basic microwave equipment operates between 7.125 Gigahertz (GHz) and 8.5 Ghz.

(1) Capabilities

The CDL is a full-duplex system that can transmit up to 960 channels of frequency division multiplex, or eight digital speech interpolation signals, combined in a 12.928 megabits-per-second (mbps) data rate. An encryption interface is available. The CDL presently transmits two-way computer data at a clock rate of 230.4 kilobits-per-second (kbps) through the DQ 11/Varian 620 interfaces. A clock rate

of 1.0 mbps is achieved through the DMR 11 interfaces. A data rate of over 200 kbps has been achieved through both interfaces.

(2) Limitations

Line-of-sight must be maintained between the microwave antennas. The CDL supports the MMCS A-Station and B-Units but they cannot be used concurrently.

4. Support Instrumentation Systems

a. Data Reduction center (DRC)

The Data Reduction Center provides the processing capability to extract data from VAX computers, personal computers, and terminals to produce reports of test results. Software packages support data base management, statistical analysis, and spreadsheet and graphical hard copy output.

(1) Capabilities

- Build and combine data bases for each field test conducted.
- Compute statistical analysis of data based on field test methodology.
- Produce statistical and graphical output for field test reports.
- Provide feedback for quality control to test data.
- Provide data fields on magnetic media for field test proponents.

(2) Limitations

- Real-time data are transferred from the ICN manually via removable RA60 disk packs.

- Video data must be analyzed, and data extracted and manually entered in the DEC VAX-8600.
- Document production is limited to line printers and X-Y plotter capabilities.

b. Range Measuring System Input/Output Test Set (RIOTS)

The RIOTS is an instrumentation tester designed by TEXCOM FHL. The RIOTS is capable of performing the identical logic functions as its RMS counterparts (Micro-B and C-Station) without requiring an RF link.

(1) Capabilities

- Player instrumentation maintenance.
- Player instrumentation countdown.
- Player instrumentation design and test.

(2) Limitations - None identified at this time.

c. Range Timing System (RTS)

The RTS provides standard time referenced to the Geostationary Orbiting Environmental Satellite (GOES). GOES transmits Universal Coordinated Time (UCT), also referenced as Greenwich Mean Time. Components of the RTS include a satellite receiver, a 10 watt, 143.20 megahertz (MHz) transmitter, and FM receivers. The transmitter is FM, and the modulation is narrow band Inter-Range Instrumentation Group, B Format (IRIG-B).

(1) Capabilities

- 125 time code receivers are available in the inventory. Ten of these are installed in panels and 11 in sync (synchronization) boxes. The remaining are available for future test needs.
- The sync boxes are used to initiate time generators in audio and video recorders. The receiver panels provide continuous time to readout devices such as Time Code generators (TCG), time code readers, and time displays.
- Five portable satellite receivers are available for test purposes or off-post requirements.

(2) Limitations

Time synchronization is presently accomplished on the players on the ground prior to start of trial. Real-time synchronization may be possible on some ground types and more possible on aircraft players.

d. Range Communications System (RCS)

The FHL RCS provides communications during tests and between tests. The RCS provides field to field, field to fixed station, and fixed station to fixed station communications.

(1) Capabilities

- Nine RCA Series 1000 repeaters, four Motorola Micor repeaters (Data Encryption Standard (DES) compatible).
- 20 Motorola MCX-100 DES transceivers and approximately 10 RCA/Tactec TAC 200 units installed in contractor vehicles.
- Two Motorola syntor at 35 W output.
- Four Regency model 250A at 25 W output.
- One Aerocom Six at 25 W output.
- One Midland 70-340A at 30 w output.

- 128 Motorola model MX-300R 12-channel DES units.
- 15 Motorola model MX-330 2-channel units.
- 14 Motorola model MX-340 12-channel units.
- 105 RCA/Tactec TAC 100 two-channel units.
- Two RCA/Tactec TAC 100 six-channel units.

(2) Limitations

- Deadspots - FHL is situated in hilly terrain which allows for permanent or variable dead spots in coverage for ground-based UHF/VHF and FM communications.
- DES Coverage - Communication range is decreased by approximately 75 percent by DES operation.
- DES Compatibility - All RCS DES equipment is per Federal Standard 1027, as supplied by Motorola Communication. All other encryption systems (such as VINSON) are not compatible with the FHL RCS DES.
- DES Radio Primary Power - Loss or interruption of primary power to mobile, base, or portable radios will cause erasure of the DES key (code) and render the units unusable for DES operation until the radio is reprogrammed with a key.

e. Voice Recording System (VRS)

The VRS records voice communications from radio, telephone, or direct input sources via magnetic tape. Time is simultaneously recorded with voice data to provide a chronological record of communication. The VRS consists of two independent recording vans and external antennas for RF reception. A data reduction subsystem is also included in VRS for data analysis of the recorded tapes.

(1) Capabilities

- Tactical FM voice communications in the frequency range of 30 to 75 MHz.
- Very High Frequency (VHF) AM (Amplitude Modulated) is receiving subsystem is capable of receiving tactical aircraft communications in the 116 to 150 MHz range.
- Ultra High Frequency (UHF) AM receiving subsystem is capable of receiving tactical aircraft communications in the 225 to 399 MHz range.
- Four, seven channel reel-to-reel tape recorders that utilize 1/2-inch-wide magnetic tape. All tape decks provide an extra tape edge track for narration purposes.
- Six Variable Gain Voice Amplifiers (VGVA) - Maintain a constant audio-level input to the tape recorders. They can also be patched in to verify data being recorded.

(2) Limitations

Data reduction and analysis is impeded due to lack of tape search and time code display capabilities.

f. Electromagnetic Interference (EMI) Van System

The EMI van system is used to perform EMI frequency surveillance and direction finding. It is also used to monitor authorized and assigned radio frequencies at FHL.

(1) Capabilities

- Surveillance frequency range of 100 Hertz (Hz) to 40 GHz.
- Automatic Direction Finder (ADF).
- Tactical communications in the 30 to 79.95 MHz range.
- Communications scanning in the VHF/UHF frequency bands.

(2) Limitations

- Manual direction finding.
- Maximum signal of 40 GHz may be detected and analyzed.

- Not being a computer-controlled unit limits data storage to hard copies.

g. Programmable Microprocessor Units

The Programmable Logic Box (PLB), the Serial Programmable Logic Box (SPLB), the Microcomputer Unit (MCU), the Serial Programmable Instrumentation Pallet System (SPIPS), and the Upgraded Logic Module (ULM) are microprocessor-based units that give each player efficient two-way data communication with the ICN for test data collection, player command, and test control.

(1) Capabilities

- Remote I/O Box and Aircraft I/O Box - 13 input event lines are available for coding in both input transition-high and input transition-low states. These events can be a primary weapon fired, trial started, video recorder on, etc. The events can also control action such as turning the laser on.
- Direct Fire Simulators (Lasers) - The PLB, SPLB, MCU, ULM, or SPIPS controls the firing rate (100 times per second), the on-time, and the inhibit-time of the pulse-position-coded infrared lasers that are used as weapon simulators.
- Hit Detectors (Laser Sensors) - A hit has occurred when a laser beam is incident on any of the laser sensors, and the PLB, SPLB, MCU, ULM, or SPIPS determines that the pulse-position-code is valid.
- Pulse Counting (Altimeters or Gun Azimuth System (GAS)) - Systems that output a frequency, such as radar altimeters, barometric altimeters, or the GAS, use the radar altimeter input of the PLB or SPLB.
- Player Alert - The player alert devices include a sonalert (located in the remote I/O box and infantry back system) and a speech synthesizer (located in the ULM Vestpack and used with 35 remote I/O boxes to put natural-voice messages onto a small speaker to alert the player of test events).

- Relay Control (Recorders or Strobe Lights) - Two relay outputs are available in the PLB, SPLB, and SPIPS for turning on recorders and strobe lights.
- Serial Bus - The SPLB uses a serial bus for communicating with other microprocessor-driven systems.

(2) Limitations

- The PLB, SPLB, MCU, and SPIPS are only compatible with the Infantry Direct Fire Simulator System (IDFSS) and the large and small gun lasers. Only the odd laser IDs can be decoded because of the cycle time of the microprocessor.
- The resolution of the time tag is 10 milliseconds.
- There are 13 line available for the coding of events to be transmitted.

h. Programmable Bus Monitor (PBM)

The PBM provides the capability to interface with and monitor 1553 multiplexed (mux) bus data. It is directly compatible with the FHL RTCA microprocessor instrumentation components and digital video encoders for post-trial analysis. The PBM can be programmed to select data from the mux bus words and to provide appropriate output signals to the FHL microprocessor elements and digital encoders.

(1) Capabilities

- MIL-STD-1553 A/B mux serial bus data.
- Input of video or audio signal.
- 64 channels for RTCA output.
- Parallel - 44-bit-wide, two channels, for video encoders.
- Serial - 8-bit, two channels, for video encoders.
- Video - three channels, buffered.
- Audio - three channels, buffered.

- Programmable.

(2) Limitations

The programmer interface connector (J14) is located on the bottom panel of the PBM. This requires that the PBM be unbolted from its mechanically mounted fixture to provide access to J14. Electrical connections must remain attached to the aircraft in order to supply power to the programmer. The PBM interface cable will provide power from the aircraft and the capability to transfer data.

i. Video System

The video system is used for recording test data from tactical vehicles, troops, weapons, and aircraft. The video system is capable of recording audio, video, and digital information. Recordings made during a test can be played back at normal speed, slow speed, or stop motion for data analysis. Time characters synchronized to the Range Timing System (RTS) are superimposed on the video screen to provide a time base for quantitative measurements. Up to four independent event markers can also be inserted in the video. Cross Hair reticule patterns can be superimposed on the video screen to provide a system boresight reference from which accurate measurements can be made directly from the video screen.

(1) Capabilities

- Video Cameras - Several different types of video cameras are available. The video cameras can be used in daylight, low visibility, and night time conditions.

- Lenses - A variety of lenses are available for a given situation.
- Video Tape Recorders (VTR) - FHL uses 3/4-inch and VHS video cassette tapes as its standard video format. There are two 3/4-inch formats in use. One is the U-matic format that allows a maximum of 30 minutes recording time at 535-line scan rates. The other 3/4-inch recording format allows recording time of one hour at 875-line scan rates. The VHS recording 1/2-inch format is generally used at a two hour recording speed but it can also be operated at a six hour recording speed with little degradation in playback quality.
- Video Monitors - A wide assortment of video monitors are available.
- Time Code Generator (TCG) - The Datum Model 9150 TCG operates as a TCG using internal 1 MHz, 5 MHz, or 10 MHz oscillator as a time base. It is synchronized to the RTS or wherever a source of IRIG-B is available. The TCG displays time in hours, minutes, seconds, and tenths of seconds. The three-digit-days display can be used to indicate event flag status of the PLB, ULM, etc.
- Cross Hair Reticle Pattern Generator - The generator inserts an adjustable horizontal and vertical bar into the video scene. A box (with size, shape, and position adjustment) is also inserted on the scene.
- Enhancer - This unit enhances the leading and trailing edges of the contour. This more clearly defines the outline of the subject.
- Time Synchronizer - Used to synchronize the internal oscillator in the VCTG/TCG to the RTS. The synchronizer, which is a portable, battery-operated unit, is also synchronized to the RTS. The time synchronizer is then used to synchronize each of the TCGs.
- Video Instrumentation Mobile System (VIMS) - The VIMS is a self-contained mobile video laboratory used primarily to support special, one-of-a-kind tests to gather information for use in making instrumentation support decisions. The VIMS is configured to allow rapid response, provide instrumentation versatility, and provide support at most locations within FHL. The system consists of local and remote video camera assemblies. The remote camera is mounted on a pan and tilt mount, and can be situated up to one mile from the control and recording center in the van. Controls for the local and remote camera include azimuth

and elevation pointing of the mount and control of focus and zoom functions.

- Video Microwave System - The system consists of two complete duplex-directional relay links or two complete omnidirectional links. The directional relay links operate at 21 to 24 GHz and have an effective line-of-sight range up to five miles. System bandwidth allows transmission of 875-line video signals with little degradation. The omnidirectional link is designed specifically for duplex audio and video transmission between a helicopter and a ground station. This system operates in the 2.3 to 2.5 GHz band and also has a bandwidth that will permit transmission of high-resolution (875-line) video.

(2) Limitations

- Scene Lighting - Camera and lens sensitivity require equipment (except low-light-level cameras) to be used in daylight.
- System Resolution - Record-to-playback resolution is limited to less than 300 television lines. The stop-motion playback is limited to one-half available resolution.
- VTCG - Video input signal must adhere to Electrical Industries Association (EIA) Standard RS-170 for 525-line scan rate and EIA Standard RS-343A for 875-line scan rate.
- Time Accuracy is limited by drift rate of internal oscillators.

j. Video Data Reduction/Debriefing Center (VDR/DC)

The VDR/DC provides facilities for post-mission playback of video recordings for data extraction and crew debriefing.

(1) Capabilities

- Playback and Monitoring Capabilities - Each VDR work station contains equipment to allow playback and monitoring of the various video formats.
- Audio and Video Switching Unit - Each workstation is equipped with a switch that allows routing of video

signals (from a playback unit) through a digital data reader and/or an X-Y coordinate measurements unit to any of the three monitors.

- X-Y Coordinate Measurements - Measurements are made from a video scene using an X-Y coordinate digitizer. This device superimposes a set of cross hairs in the video scene and digitizes their location. Position of a target with reference to video camera boresight can readily be measured with the digitizer cross hairs. Typically, the system can produce angular measurements to any accuracy of 0.1 mrad.
- Digital Data Extraction - Provides the capability to capture, record, and play back digital information superimposed on a video signal. This capability was originally developed to capture and record information from the MIL-STD-1553 data bus; however, with suitable interfaces, the system can record any type of digital information.

(2) Limitations

Search rate for the two VHS playback units is limited to two times normal speed. It is highly desirable to be able to search video tape at least 10 times normal speed. Improvement programs to accomplish this have been identified.

k. Automatic Direction Finder (ADF)

The King Radio KR87 ADF is a digitally tuned, solid-state receiver that provides bearing information in the 200 KHz frequency band. The unit's gas discharge display always displays the active ADF frequency in the left window. The right window displays either the standby frequency, which can be transferred to become the active frequency, or a unique flight timer or programmable elapsed timer.

(1) Capabilities

- The ADF automatically measures bearing with respect to a frequency source (Amplitude Modulation (AM) radio station or beacon).
- Under ideal conditions, the ADF is accurate to three degrees when the radio frequency strength is at least 70 microvolts per meter.
- Accuracies on land vehicles vary from +/- 10 to +/- 25 degrees, depending on the vehicle and antenna location.

(2) Limitations

- Full ADF accuracy is not achieved on some of the vehicles because the antenna cannot always be mounted in the best position on the vehicle.
- Tactical radio transmissions interfere with the ADF.

1. Strobe Light Assemblies

The Heathkit and Hoskins aircraft strobe light assemblies are self-contained strobe lights mounted on ground vehicles. The strobe lights are modified to allow remote control from the central computer through the Programmable Logic Box (PLB) or Serial Programmable Logic Box (SPLB).

(1) Capabilities

- The Heathkit strobe light assembly provides the primary kill cue notification to the ground vehicle personnel. The strobe light flashes continuously at approximately 60 pulses per minute when activated by the computer control relay.
- The Hoskins strobe light assembly is used as a kill indicator, a pairing indicator, and/or a weapons simulator. The strobe light flashes one time for each activation of the computer control relay.

(2) Limitations

- The Heathkit strobe light cannot be used for pairing of fire indicator because the electronics do not allow coding of flash at different rates.
- The strobe lights should not be powered from nickel cadmium RACAL/BA-4386 battery sources due to high power drain.

m. Instrumentation Power System (IPS)

The IPS removes the harsh, ripple, dips, and transients from vehicle power. This allows vehicles to power sensitive instrumentation and to charge the instrumentation batteries. The IPS provides unregulated 24 volts, Direct Current (VDC), regulated 13.8 VDC, a vehicle power ON light, and a low voltage warning light with corresponding logic signal for Built-In-Test (BIT)/Computer-Aided-Test (CAT).

The IPS consists of six basic elements:

- Transient Suppressor - Accepts the noisy input from the vehicle power bus and, by means of a multi-state L/C filter, reduces the worst-case transients by 1000:1.
- Reverse Blocking and Charge Limiting - The instrumentation batteries are protected from discharging back through the vehicle's circuits by a 50 A (ampere) Schottky diode.
- Low Impedance Filter - A 65,000 microfarad capacitor on the output flattens vehicle generator ripple and minimizes instrumentation switching transients.
- 12.0 V Regulator - A voltage regulator provides up to 4 A at a fixed 12.0 V as an alternate output.
- Low Voltage Warning - Uses an op-amp (operational amplifier) with 100 millivolt (mV) hysteresis to provide snap action and prevent chatter.
- Instrumentation Batteries (External to IPS) - The instrumentation batteries are included because they are the major source for instrumentation power when the

vehicle engine is not running and the sole source for instrumentation power when vehicle power is turned off.

(1) Capabilities

- Test data on vehicles show that IPS reduces transients from 20 V peak-to-peak (p-p) to 0.2 V p-p at 1.5 A load (and to 0.1 V p-p at 1.0 A load).
- Very little current (58 milliamperes) is drawn from the vehicle batteries by IPS with the vehicle master switch on.
- Six parallel connectors provide parallel outputs to handle the multiple requirements of instrumentation.

(2) Limitations

Certain precautions for cabling and weatherproofing the IPS as outlined in the IPS Operation and Maintenance (O&M) manual must be observed. IPS is designed to carry 30 A continuously and is limited by a 40 A fuse on the input. When connected through the MX-7777 Transient Suppressor, the 50 A circuit breaker will provide adequate protection against high current shorts.

n. Engagement Line-Of-Sight System (ELOSS)

The ELOSS identifies periods of time when members of opposing forces are not directly separated by terrain, i.e., when masked. Masking and unmasking events are transmitted through the RMS to the MCS. The state of being unmasked indicates that direct fire is no longer precluded by intervening hills, trees, bushes, etc. In simulated direct fire engagements, the time interval between exposure and

laying fire is a performance measure of target finding, acquisition, and recognition systems.

The ELOSS simulates visible light with invisible and omnidirectional RF transmissions between interrogators and transponders mounted on opposing vehicles. Both interrogators and transponders are essentially RF beacons. Two-way transmissions and receptions that are unattenuated by intervening terrain are taken to indicate mutual exposure. The absence of completed transmission and receptions by the interrogators above threshold power level are taken to mean that terrain is providing a mask.

;(1) Capabilities

- Range - The range of operation between interrogators and transponders is 500 m to 15 km.
- Playing Time Considerations - The interrogators require a line-of-sight with the master clock for initial synchronization, after which these units maintain synchronization for at least two hours. Resynchronization occurs automatically whenever the master clock pulse is received.
- Polling Rate - The polling rate is once per 750 milliseconds (ms) with the maximum of 25 interrogators. The polling rate is established by the number of interrogators used and is set and controlled by the master clock. The vehicle instrumentation checks the interrogator output for masked/unmasked state changes every 0.5 seconds. Individual state changes are ignored. Two consecutive outputs of a changed state are accepted as an actual change and are transmitted in the RMS as such.

(2) Limitations

- Numerous defilade test have shown that correct intervisibility decisions are made at distances greater than 0.5 m, above and below grazing where the edge of the mask is in line with the two players.
- At this point in time, ELOSS has been verified to 10 km through the use of video and post-trial real-time data reduction. To validate ELOSS out to the range of 15 km will require further testing with more sophisticated video equipment than is currently available at FHL.
- Upgrading of the ELOSS system is partially complete. The solid-state transponders and the field test sets are part of this upgrade. The upgrade will be complete when solid-state interrogators are procured, tested, and operational.

o. Laser Spotting Information System (LSIS)

The LSIS was designed to record digital and video data on the return energy of laser designators employed in military weapon systems. The LSIS is a field deployed system that collects digital data on the amplitude and time shift of the return laser energy and determines the range to the targets. The video data collected shows the exact location of the laser spot on the target in real-time mode. The LSIS has three main subsystems: video, radiometric, and data reduction. Additionally, three RF systems, a microwave and two VHF, are used in remote deployments.

(1) Capabilities

- Video - The video system provides the position of the laser spot on the target. Two cameras, both RCA TC2001, are mixed together to provide spot video. A 2,000 millimeter (mm) Celestron C-8 lens provides for ranges up to four km. A third RCA camera (wide angle) with a 150 mm lens is boresighted to the assembly to enable the tracker to find targets in a wide field-of-view.

- Radiometric - The standard configuration involves the use of one LFD and one radiometer. The LFD is pointed at the designator and picks up the lasing start pulse and an analog pulse. The analog pulse is used to blank the equipment to enable collection of return pulses.
- Data Reduction - The video tapes, recorded during previous missions, are threaded onto the Logicon tape drive. Encoded information is in EBCDIC (Extended Binary Code Decimal Interchange Code) format, and the DEC equipment operates in ASCII (American Standard Code for Information Interchange).

(2) Limitations

Proper data collection from the LSIS can only be guaranteed if the system is deployed in a rigidly structured fashion. The system cannot support freeplay scenarios. The location of the designator and the targets must be known at all times, and the targets must travel predetermined routes. Additionally, there are limitations on the speed with which the players can travel. It is suggested that practice runs be made prior to record trials to ensure trackability.

E. FUTURE INSTRUMENTATION

The TEC instrumentation system has competently served the cause of testing well over the years. Numerous important equipment decisions have been made using the data it produces. However, the electronics industry and the capabilities of military systems have advanced significantly since the instrumentation system was designed. It is increasingly difficult to find replacement parts for its components and

many of the weapons systems the Army is developing today cannot be simulated by this instrumentation. These limitations also apply to TEXCOM's other test agencies and several development efforts are underway to correct them.

1. Mobile Army Field Instrumentation System (MAFIS)

More than eleven years ago, a need to find a method of bringing the test to the troops was identified. This was an attempt to alleviate the social hardships experienced by the soldiers taking part in the tests and to reduce the costs of the tests. The Mobile Army Field Instrumentation system was conceived to meet this need. Unfortunately, development of this instrumentation system has produced no usable test set to date. TEXCOM recently organized a dedicated MAFIS development activity and refuted the project. The project was renamed the Mobile Automated Instrumentation Suite (MAIS) and was turned over to the Army Materiel Command (AMC) for development.

There is little hope of early success in producing a working mobile instrumentation system. The Army material command is responsible for developing multi-billion dollar weapons systems for the Army. A test instrumentation system is not likely to have a high priority when competing with these major projects.

A major drawback to developing an entirely new instrumentation system is that no vehicle exists to test

components realistically until the entire system is built. This situation can be compared to writing a complex computer program without testing its subroutines. The TEC has demonstrated the ability to integrate and test components of a new system in its RMS. It also had extensive experience with instrumentation development. It would seem judicious for TEXCOM to assign TEC responsibility for MAIS development.

2. Global Positioning System (GPS)

Use of GPS is an essential element of the MAIS concept. The TEC conducted an experiment in December 1988, which was successful in producing accurate position location (PL) data using commercially available GPS receivers. The PL data was successfully integrated into the RMS.

Position location accuracy in X, Y, and Z of one meter with GPS is possible. When combined with a Navigation Integrated system (NIS), accuracies of .1 degree in heading and .3 degrees in roll and pitch are also possible. This combination also provides accurate PL as the GPS receiver moves in and out of view of the satellites in wooded terrain. The current POSNAV (Position Navigation) system being developed for the M1A1 tank could provide the NIS component of the system. The TEC is designing a new player instrumentation subsystem, the vehicle logic module, in a way that allows POSNAV and GPS to be integrated into it.

The limited ability of the RMS to move data (approximately 20,000 bits-per-second) complicates test planning at TEC. The GPS and other new instrumentation will significantly increase the amount of data produced. The TEC now transmits much of its test data in video or manual form because the RMS cannot handle more data to speed up the process. An improved data transmission system to move the information generated by this system must be developed. Efforts are currently underway at TEC to develop improved transmissions of data.

3. K-Band TRADOC Obscuration Pairing System (K-TOPS)

The K-TOPS is being developed to simulate the ability of weapons to sight and fire through obscuration. This system is based on police-type radar technology. It is used to allow weapons pairing through levels of smoke and fog that defeat current laser-based systems.

4. Non-Line-of-Sight Testing

Realistic simulation of non-line-of-sight weapons is one of the biggest challenges facing operational testers. In 1987, TEC organized a special projects office to develop tools needed to meet this challenge. The TEC approach envisions establishing a one meter accuracy database of the terrain in its primary test laboratory at FHL. This involves assembling the hardware and writing the software to digitize this terrain from aerial photographs. This database will then be used to

project aspect views of the terrain over which the missile is notionally flying on to the gunner's console screen. The system will be linked into the RMS, which will insert realistic icons representing other players on the terrain in what would be their actual position if the missile were flying.

While the simulator will allow TEC to test the effect of non-line-of-sight weapons in battle, it will not elicit realistic reactions from the target crews. This necessitates the need for a surrogate missile. In partnership with the Navy at its China Lake testing area, the same special project group is modifying a small drone aircraft to fill this role. The drone is launched and held in a loiter area by its operator until the non-line-of-sight gunner fires a missile. At that time, the drone aircraft flies toward the target, controlled by commands from the gunner's console. When the gunner starts the missile on its final attack, the drone breaks off its flight and returns to the control of its operator for recovery. A tape of target impact is projected on the gunner's screen.

F. Summary

The most important aspect of OT&E is that it provides an independent evaluation of the utility of a system. Equally important, it also provides an evaluation of the feasibility of employing a system. It is crucial that the testing

community provide this system evaluation in an efficient and competent manner.

III. PROBLEMS/CRITICAL QUESTIONS RAISED BY PROGRAM MANAGERS

A. INTRODUCTION

This chapter focuses on providing answers to the most frequent/critical questions asked by program offices concerning current operational testing [Ref 9]. These questions include: "What type of instrumentation exists to accurately represent and test my system? (answered in Chapter II), "Why do testers need access to the development system so far in advance of the test?", "Why is there so much preparation prior to conducting a trial and why can't it be done faster?", and "When will the test report be finished?" These questions will be answered in this chapter using the SGT York (DIVAD (Division Air Defense)) test, conducted at FHL, as an example. However, one tool, Model-Test-Model, that has been developed with the goal of optimizing testing and is part of the answers to the above questions will be covered in detail in Chapter IV.

B. WHY TESTERS NEED ACCESS TO DEVELOPMENT SYSTEMS EARLY IN THE ACQUISITION CYCLE

"First article" systems refers to a representative system that will be instrumented during a field test. This term not only refers to the system being evaluated but also to any

system that will partake in the field test. The SGT York test is a good example of this situation. The SGT York (DIVAD) antiaircraft system was the primary system being evaluated. However, there were other systems such as the M1A1 Main Battle Tank, the Bradley Infantry Vehicle, M113 Armored Personnel Carriers (APC), M60 Tanks, the Stinger air defense system, the Chaparral air defense system, an Air Force F-111 aircraft and A-10 Thunderbolt aircraft, a Navy A-4 aircraft, and AH-1S Cobra helicopters. Every one of these systems had to be instrumented to enable the SGT York to be evaluated. This was a daunting challenge for the testers at FHL since some of these systems had not been instrumented before by FHL.

I was a Project Engineer assigned to FHL during this test and was one of the primary people responsible for the instrumentation packages. Since we were to instrument some systems we hadn't instrumented before, we needed them as far in advance of the scheduled date of the test as possible. Even with the systems we had previously instrumented (such as the Chaparral and AH-1S Cobra, M60 Tank, and M113 APC), we still needed to check them to ensure that modifications had not been made to the systems. To appreciate why we as testers fight so hard to obtain first article systems, some systems and the problems we encountered will be detailed.

1. SGT York

The SGT York provided us with some unique challenges. The SGT York was equipped with an acquisition and tracking radar and twin 40-mm guns. Once the acquisition radar acquired a target, it would hand off its information (velocity, range, etc) to the tracking radar and then the lead and elevation angles for the guns to fire would be computed. This would allow the SGT York to shoot a volley of shells into the path of the aircraft and the aircraft would fly into them and be destroyed. This presented us a problem because we use lasers to simulate the firing of a weapon. Unfortunately, the lasers need to be pointed in the direction of the target in order to register a kill (the lasers simulate a direct fire weapon). This was not possible with the SGT York since its twin 40 mm guns never point directly at the target. A solution involved using a unique device to monitor the 1553 data bus and extracting the necessary bits of information to relay them to the FHL computer complex. The computer would then determine if any aircraft were in the vicinity of the firing SGT York and determine the probability of kill. As luck would have it, the device we needed was not on the market and had to be designed, built and tested. Not only did this require outside contractors, but also FHL software designers had to work many manhours of overtime to develop an algorithm for this computation.

2. F-111, A-10, and A-4 Aircraft

These aircraft presented not only instrumentation challenges but coordination problems as well between FHL, and the participating Air Force and Navy units. The aircraft presented two problems: 1) An instrumentation pod had to be designed and built and 2) the instrumentation power provided by the aircraft had to be heavily filtered so that the instrumentation would not be ruined. Once the pods were designed and built, the FAA (Federal Aviation Agency) had to certify the airworthiness of the pods. Anyone familiar with bureaucracies can appreciate how long this certification took. The FAA took approximately 3 months to officially certify the pods and this required them to do some exceptions to policy to get it done that "fast".

3. Chaparral

The Chaparral presented a problem that involved connectors. Although FHL had previously instrumented the Chaparral in years past, we still had to build some interface cables. We did not have any of the mating connectors for the interface cables in stock. When we tried to procure the particular connector we needed for the cables, we found that no manufacturer currently built it. Therefore, we had to pay substantial retooling and production costs for a limited run. This process took approximately 4 months from processing of the purchase order to receipt of the connectors.

These sample problems that occurred during the SGT York test all required weeks and months to resolve. Confounding the problem is trying to use the Army/DOD procurement system which is not "user friendly". The procurement system is not set up to procure items that are needed expeditiously. Obtaining the needed item can take anywhere from a month to a year.

The SGT York test was used as an example to answer the question of "Why do the operational testers need the development systems so far in advance of the test?" Program offices need to understand that they cannot bring their systems to the testing facility a few days before the test date and expect the test to go as planned. Having testers involved at the beginning of a program and allowing them early access to systems can only benefit the system and the program offices.

C. WHY TEST TRIALS TAKE AS LONG AS THEY DO

Test trials commence once the instrumentation packages for each system have been designed, built and tested. On the day of the scheduled tests, the systems are moved to the staging area. The staging area is where all the systems (players) have each piece of equipment in their instrumentation packages checked prior to a test trial. This pre-trial equipment check is crucial to ensuring that a valid (nonbiased) trial is conducted. This equipment checkout is also commonly known as

an instrumentation "countdown". The time to complete this procedure depends on the number of players and on the complexity of the instrumentation packages. For the SGT York and AHIP (Army Helicopter Improvement Program) tests, the countdown would start at two a.m. for a trial beginning at eight a.m. (FHL is continually trying to improve its testing and have been able to speed up this process [Ref. 10]). Players will be not be allowed to leave the staging area unless their equipment is operating properly.

Why is the countdown so crucial to a trial? Testers are responsible for ensuring that if a system is shown to have deficiencies it is not due to the test equipment. The purpose of testing agencies is not to make judgements on the system but to provide information about it. Therefore, testers are continually concerned about introducing bias into a test.

D. THE FINAL TEST REPORT

Operational test reports have to be considered by program offices early in the acquisition cycle. This is due to many program offices being dumbfounded over how long it takes for a final test report to be produced. As more and more requirements are proposed, then more trials will have to be conducted, more data gathered, and more data reduction and analysis will have to be accomplished. It is not uncommon for a final test report to be published up to six months after a major test.

One of the major contributors to data reduction time is the increased use of video to gather data. Video is typically used to monitor crew operations and workload. The advantages of using video include being able to monitor what is actually taking place. The disadvantage of using video is that reducing the video data into measurable parameters is very time consuming.

Other data parameters are sent to the main processing computer where initial data reduction is performed. The initial analysis determines whether the completed trial was valid (no bias introduced into the test or no major equipment outages). Once this initial analysis is conducted and the trial is deemed valid, then the test data is stored in the main database for further reduction at the conclusion of the test. The TEC is currently in the process of upgrading its computer network and should be able to process data at a much greater capacity.

Interim reports have typically been issued to interested agencies in the past. The purpose was to provide some type of early analysis to the interested agencies. However, these interim reports had to be viewed with extreme caution. Program offices need to be aware of the data reduction requirements and be ready to augment analysis agencies such as AMSAA (Army Material Systems Analysis Agency) if necessary to expedite a test report. Program offices also need to be aware of tools such as Model-Test-Model that has the potential of

keeping trials and players at a minimum to answer the operational issues of a particular system. This would greatly reduce the time to produce a final test report.

IV. MODEL-TEST-MODEL

A. MODELING INTRODUCTION/BACKGROUND

Military organizations have been the source of much of the development of modern, sophisticated modeling technique. World War II can be taken as an arbitrary historical starting point and perhaps specifically the early British "operational research" on problems such as the operational use of radar and the design of aircraft search patterns for submarines in the North Atlantic. But the concept of models and modeling is neither new nor specific to military applications.

The concept of a model is very broad and general, but always subject to constraints. A model is potentially useful to analysts and decision-makers because it represents the real world but does not replicate it. Replicating the complex real world would neither be feasible nor desirable. One could spend eternity in futile attempts to take account of all the possible variations that the complex world makes possible, in nonsolutions or even in intelligent but unacceptably slow experience gathering and untested learning. Rather, it is better to simplify particular aspects of the real world to help in solving particular problems. One doesn't want to try to represent everything in an all-purpose model that tells

everybody everything, solves nothing, and takes forever doing so.

1. PURPOSES OF MODELING

Models cannot always solve problems. Many problems cannot be solved, particularly in the military field in which answers can only be determined in war. But when models cannot provide solutions they should always at least "shed light." They may do this in several ways:

The process of constructing and using a model should increase the understanding by both the analyst and his client of the process or problem being studied. It can be heuristic, helping the analyst to find ways to point to a solution. But the purpose is not just to educate the modeler. The learning must be transferred to the user. Communication and interaction between modeler and user must be continuous and open.

Models can aid in making choices even in situations of high uncertainty. Doing nothing is choosing one alternative. Models can assist in comparing alternative weapon systems, tactics, environments, routings, and training methods. Relative numbers are what count in selecting among alternative means.

Models may sometimes give answers in the sense that the absolute numbers are taken as valid. A limited logistics

model may be able to give estimates of absolute quantities of fuel consumed, or vehicles required in given circumstances. However, caution must be used when dealing with absolute numbers as the following passage illustrates.

A bomber penetration model may give 70 percent bomber survival, or 70 percent of targets hit. We cannot know that 70 percent would be the real number if there were a war, even though we may have high confidence that the particular laydown that gave this figure will do better in a real war than an alternate laydown that gave 50 percent vice the above 70 percent when used in the same model [Ref 11:p. 7].

The purpose of a model should always be subsidiary to the purpose of the modeler or of the decision-maker he serves. Models do not analyze anything. **Analysts** analyze, and models can assist them in their task. Models should always come after the definition of the problem. Modeling is only one aid to the analysis. It is never clear that a numerical, mathematical or computer model should be used, or that a particular type of model should be used.

2. TYPES OF MODELS

There are a myriad of different types of models. These models can be categorized in many ways. For purposes of illustration, four taxonomies are considered:

- Application.
- Objective function: Effectiveness, Cost, Cost Effectiveness.
- Level.

- Technique.

- a. Application*

Application in this thesis is limited to military applications with an emphasis on Army concerns (which, however, interface in important ways with Air Force and Navy forces and operations). Models may find uses in virtually every Army activity, at every level. An example of the breadth of applications is given by the following partial listing:

- Combat Support Services (C³I, mobility, etc).
- Costing.
- Force Sizing.
- Force Structure.
- Procurement.
- Program Management.
- Strategic and Tactical Combat Operations.
- Support.
- Testing and Evaluation.

Though applicable, models may not be equally useful in all of the above categories. A given model may also involve more than one of the above applications.

- b. Objective Function*

The object (characteristic to be analyzed) may be classified in many ways. One of the most basic breakdowns is:

- Effectiveness.

- Cost.
- Cost Effectiveness.

These categories may be broken down in various ways.

Effectiveness is a term most often used with weapon systems; it may also apply to support systems and to various operations. Its meaning varies with the level of operations, or the scope of the model, and with the Figure(s) Of Merit (FOM) chosen for its measurement. Ideally, a single FOM would be the optimal situation-targets destroyed by a missile, aircraft kills in air-to-air combat, or aircraft kills by air defense. However, generally no single FOM is adequate. One has to take into account collateral damage, damage expectancy, and non-target destruction. Kills in situations such as air combat cannot ignore losses. But relative attrition rates may not tell us whether the kills were sufficient to meet the real objective: preventing the enemy from carrying out the missions of his aircraft other than interdiction or close air support [Ref 7:p. 8].

Trying to establish a single figure of merit is not the only difficulty in measuring effectiveness. For example, if we are dealing with weapons systems then we may have to consider whether we can find rational ways to compare three different categories of weapons:

- Existing weapons.
- New weapons with greater capability than existing weapons.
- New weapon concepts that do things that can't be done now.

Such dissimilar weapons are not easy to compare with each other. The problem is further exacerbated by the great difference in the time at which different systems may become available. This often means tradeoffs between better performance late and possible military risks earlier.

One must also ask whether a qualitative improvement really increases effectiveness or is simply "gold-plating". As in the case of models to which too much may be added simply because "we can do it, and someone might use it," in weapon systems some added features may even lessen effectiveness. This may happen, if, for example, the added features make an operation more difficult or maintenance more frequent. Effectiveness is also intimately tied to cost and trade-offs have to be made when dealing with budgets.

Cost can generally be estimated with less uncertainty than effectiveness, and costing is at least conceptually simpler than effectiveness analysis. This is not to say that cost estimation is not a highly complex problem in itself. The difficult problem in costing is the handling of time. This factor is implicit in the basic classification of the costs of military systems into Research & Development (R&D), Investment (Procurement), and Operating costs.

R&D is the first cost incurred for a weapon system. It may start several years before the Investment phase, although it also generally overlaps the latter, as testing

plus R&D on improvements and modifications continue. R&D is conceptually a one-time investment cost.

Investment covers the procurement of the system itself, including initial spares. It includes items with a long life, measured in years. Investment costs may also continue after full deployment is reached, as modifications of components are retrofitted to the original equipment.

Operating costs, maintenance, training, personnel costs, fuel, etc, are incurred during the operation of the system. Where there are many units of a system, operating costs start when the first unit is deployed and build up until the earliest units start phasing out. Therefore, operating costs overlap investment and sometimes R&D costs.

These three cost elements are frequently referred to as Life-Cycle Costs (LCC). This is to emphasize the importance for both analytical and planning purposes of estimating the operating costs over the lifetime of the system, especially when considering possible parametric variations in levels of procurement, rate of build-up, etc. LCC should also include estimates of salvage value.

The term cost-effectiveness was first used in post-WW II military operations research. This was in recognition of the fact that military effectiveness cannot be measured in dollars. So, it is necessary to measure the effectiveness that can be achieved for a given cost.

It is not possible conceptually, mathematically or practically to simultaneously maximize one variable, effectiveness, and minimize another, cost [Ref. 7:p. 18]. It is always necessary to fix one and vary the other.

c. Level

The scope of a model may be expressed as the level of optimization attempted. In most cases this is dictated by the responsibility of the analyst, his superior, or the unit to which he is attached. Air combat models offer us a useful "model" of a way to categorize models in ascending order of complexity and of potential learning about the processes involved [Ref. 7:p. 22].

One-on-one encounters are the heart of the air combat problem. Their outcomes depend on such probabilities as acquisition and maneuver into favorable position - adding up to the probability of kill by one or the other aircraft and reflecting both chance and what is known of the characteristics of the pilot, the aircraft, and the air-to-air missiles involved on both sides. One-on-one models may pit similar aircraft against each other, or one side may have a superior system. However, one-on-one may not be realistic in an engagement scenario.

Few-on-one may have an aircraft in combat with one or more added enemy aircraft. What are his chances of escape or of achieving one or more kills? Are there real world

limits on how long a pilot would actually fight? His superior characteristics might make excessive kills an artifact of the model, without the introduction of a realistic engagement time constraint.

Few-on-few may increase the realism. Now that we may be concerned with target assignment on each side, how do the subsumed one-on-one and few-on-one engagements add up? How do the two groups break off? Few-on-few may also involve interactions of dissimilar systems.

Many-on-many reflects a further step in realism which must include all of the above, with rules for combining these individual and group engagements. Many-on-many is also qualitatively different in that command decision must be modeled. A difficulty in the many-on-many is that sometimes the command decision model is inherent in the mathematics used and may be unrecognized by the analyst.

At the Theater level, we may not be satisfied in modeling who wins each engagement. We need to know how outcomes affect the theater war. Outcomes that need attention are areas such as interaction with the ground forces, movement of the FEBA (Forward Edge of Battle Area), and other measures of progress toward victory or defeat. Theater models are the most challenging of all, and no fully satisfactory theater model has yet been constructed [Ref. 7:p. 24].

d. Technique

The techniques of modeling are numerous and are constantly growing in numbers and sophistication. A useful way of classifying techniques was suggested by L. J. Low [Ref. 12:p. 10]:

- Analytic games - Consists of problem solving type games.
- Computer games - Consists of arcade-type games.
- Interactive computer games - Lets the players provide input into the game.
- Computer-assisted war games - Consists of Corps Post Exercises (CPX) and Field Training Exercises (FTX) where a computer acts as the umpire to determine the outcomes of various battle scenarios.
- Manual war games - Sand models, maps, and mock-ups are used with individuals acting as the umpires to determine the outcomes of battle scenarios.
- Military exercises - Actual troop exercises are conducted.

This classification by technique can be crossed with the earlier classifications in a multi-dimensional taxonomy of models. Not all intersections, or cells, would be of equal interest. For instance, one would not use base force planning on low-level (e.g., one-on-one) analytical models. Conversely, a theater or global model is unlikely to solve a problem of personnel training or fuel consumption.

e. Conclusion

These several ways of classifying models have been offered as an introductory framework into the use of models in

general. The rest of the chapter is devoted to the Model-Test-Model concept.

B. MODEL-TEST-MODEL

The Model-Test-Model (M-T-M) concept is designed to capitalize on both combat simulation modeling and field testing capabilities of U.S. Army analysis agencies. The M-T-M concept is one method that the Army can use in determining how to modernize in this era of diminishing resources. The M-T-M concept consists of five phases: 1) long-term planning phase; 2) pretest modeling phase; 3) field test phase; 4) post-test modeling and calibration phase and 5) accreditation phase. A general overview of each of the phases will be given.

1. Phase 1: Long-Term Planning Phase

The long-term planning phase begins by establishing agreements between the organizations conducting the M-T-M. The agreement is usually formalized in a Memorandum of Agreement (MOA). The MOA assigns responsibilities of the organizations involved, their working relationships and resourcing commitments, the terms under which resources will be provided and obligated, and the products expected to be delivered by each agency [Ref. 1:p. 22]. This allows for the selection of specific tests for conducting M-T-M. Because each test on which M-T-M could be conducted will vary in terms of responsibilities and resource commitments, each M-T-M

project should be initiated and governed by a separate Project Coordination Sheet by the participating elements of the respective organizations.

2. Phase II: Pretest Modeling

This phase uses high-resolution combat simulation models to help plan field tests. There are two types of pretest modeling. The first is pretest modeling in support of Force Development Test & Evaluation (FDTE). The second is pretest modeling in support of operational testing. In the pretest modeling phase, it is important to use an interactive combat simulation model because the personnel involved in developing scenarios are not usually experienced in using simulation models and interactive simulations are more "user friendly" than non-interactive models. This greatly lessens the amount of time spent on teaching personnel how to use the model. Additionally, minimal programming support is needed in interactive models when changes need to be made (scenario changes, etc).

a. Pretest FDTE Modeling

There are two distinct advantages of using M-T-M prior to the FDTE. First, the maneuver unit leaders are able to use an interactive simulation model to develop scenarios they can use during the field test. The leaders try and determine optimum tactics for given scenarios. The scenarios developed should resemble the restrictions developed by the

testing personnel. Second, results from the scenario development by the leaders can be given to the proponent of the weapon system (such as the Program Management Office) to improve their tactics prior to the test. This provides a doctrinal review of the scenarios by the proponent subject matter experts.

The primary objective of pretest FDTE modeling is to refine test design with the goal of saving time and test resources [Ref. 8:p. 277]. In doing so, it also provides training for the maneuver unit leaders and can decrease the learning curve effects during the field test. Past field tests have shown that units typically need a couple of trials for the player units to get organized.

b. Pretest Operational Modeling

As with the FDTE, an interactive model is used prior to the actual operational test. Proponent scenario and subject matter experts are consulted to develop scenarios to be used during the operational force-on-force field test. The scenarios developed should resemble the restrictions developed by the test design team. The objective prior to the operational test is to examine whether the test objectives can be met within the parameters established by the test design. For example, does the terrain where the operational test is to be conducted allow engagements by the weapon systems throughout the required ranges? The goal is to provide

information to improve test design (such as whether to focus the test in MOPP (Mission Operational Protective Posture) 0/4, defense/offense, or night/day).

3. Phase III: Field Test Phase

Once the early FDTE test have been conducted, then the system is evaluated more realistically via field testing (operational testing). These tests are conducted to assess how the system characteristics perform in various situations. Field testing is performed by military personnel in a series of trials conducted to replicate realistic employment of the system to determine lethality, fightability, survivability, and suitability.

When the field tests are being performed, some of the trials should be devoted to replicate some of the successful tactics developed by the maneuver unit leaders during the pretest modeling. This allows the maneuver units to test some of the tactics that were successful in the combat model simulations. It also enables the maneuver units to see if the pretest model tactics hold up under field conditions.

During operational testing, tactics must be unscripted (leave as much randomness in the test as possible), but must remain within the limits established by the test design (i.e., terrain boundaries, force size, etc). Past efforts in M-T-M resulted in the modeler expressing a desire to script the operational trials in order to better establish a basis for

model validation [Ref. 8:p. 278]. However, this philosophy defeats a primary tenet of operational testing that test players need to be free to use tactics that they would realistically employ on the battlefield. This provides more credible use of the system being tested.

Other than scripting the test, it is essential that the modeler be involved in the field test to better understand the test conduct and data. He/she should be present at the test site observing trials and also coordinating with the data reduction group to better understand the format of the test data. Once the field tests are culminated, the data is then passed on to the modeler to begin post-test modeling.

4. Phase IV: Post-Test Modeling and Calibration Phase

During post-test modeling, the objective is to refine or calibrate the simulation/model in preparation for validation. This phase begins with the evaluator selecting the trials to be used for post-test modeling. The objective is to have the selected trials replicate the field test as close as possible for validation purposes. This requires that model input parameters be determined and updated if the same model is used from the pretest modeling phase. The same procedure must be used for the field test constraints that are input into the model.

The next step involves determining what measures of performance to use for comparing the field test data to the

model output data. The comparison can be accomplished using the Micro/Macro approach. The micro approach of the model/test comparison is the primary approach recommended [Ref. 8:p. 279]. If this method is used the model must also replicate actual player states (i.e., routes, orientations of weapon systems, etc). The micro comparison analysis is conducted at the individual events level (i.e., detections, engagements, and player movement rates). Parametric and nonparametric goodness-of-fit tests on distributions can be used to make the final comparison.

The alternative macro approach for model/test comparison is least preferred. The reason is that the macro measures could be the same for the model and field test data, but much could be happening in the model that completely differs from individual events. For a particular trial, input the number and type of players from the test into the model and run iterations. Compare the model output to the field test using macro measures.

There will never be an exact match between the model and the field test. Closer agreements can be obtained through calibration of the model to the test. The following areas need to be rechecked for causes of model-test differences: data from the field test, model input parameters, and algorithms in the model and test. Changes can then be made to modify a test version of the model to replicate the field test within tolerances. Any major changes to the model need to be

provided to the organization that holds configuration control for proposed changes to the reference version of the model.

5. Phase V: Model Validation

Model validation will continue to be a key area in M-T-M. Before a combat simulation model can be used to extend the M-T-M test results, it must be validated or at least accredited. This is the most difficult task for the modeler. The modeler must prove the credibility of a simulation model to the user. "To say the simulation model results are credible implies evidence that the correspondence between the real world and the simulation is reasonably satisfactory for the intended use" [Ref. 12:p. 13] (i.e., the model is validated for testing and evaluation purposes).

In accomplishing the final phase, the M-T-M team puts together an accreditation package that consists of the results of the model/test comparison analysis conducted during post-test modeling. The package serves as the credibility documents for the user. The package is reviewed by the end user of M-T-M (such as OPTEC). OPTEC (Operational Test and Evaluation Command) makes an accreditation decision on the model for testing and evaluation purposes. Once the model is accredited, then the model can be used at OPTEC's discretion to extend test results beyond the test environment.

The test constraints that were input into the model during the pretest and post-test phases can now be removed.

The doctrine or system being tested can now be carefully evaluated in a different environment. Different types of terrain, weather, and force size can be used in the model to see how the doctrine or system performs. A weapon system may appear to meet performance standards in Fort Hunter Liggett's terrain with a limited size force, but not perform to standards in the Middle East with a larger size force. The goals for post-test modeling are to improve the model, validate the model, and use it for cautious interpolation and extrapolation.

C. Model-Test-Model Using Janus

This section deals with the actual use of the M-T-M concept (post-test phase) using Janus to simulate events that occurred during a field test conducted by TEXCOM Experimentation Center (TEC) at Fort Hunter Liggett, California. The results of the field test were entered into the Janus model and the results were documented in a research paper by TRAC-Monterey. The research paper is entitled "Comparison Of MIA1/M1A2 Battle Results Between Janus(A) and an Operational Field Test" [Ref. 16:p. 1].

1. Janus Overview

Janus is an interactive, stochastic, two-sided, force-on-force, high-resolution combat model used extensively throughout the Army. The original version of Janus was developed at the Conflict Simulation Center at Lawrence

Livermore Laboratory. It was later improved (Janus (A)) by the Janus Working Group at TRADOC Analysis Command (TRAC), White Sands Missile Range.

Janus (A) models individual level systems, such as tanks, helicopters and soldiers, the battlefield environment, and each system's interaction with other systems and their environments. The characteristics of these combat systems include descriptions of the weapons carried, weapon capabilities, etc. In addition to modeling individual systems, Janus (A) is also capable of modeling aggregate forces.

2. Comparison of M1A1/M1A2 Battle Results Between Janus (A) and an Operational Field Test

a. Introduction

The purpose of this modeling effort was to report the post-test modeling results from the comparison of field test data from the M1A2 Early User Test and Experimentation (EUTE) with data generated from modeling of those field trials in Janus (A). The modeling effort had two primary goals. The first goal was to attempt to accredit Janus (A) for use with M-T-M, specifically in its representation of the M1A2 tank. The second goal was to identify or verify Janus (A) modeling shortcomings and recommend improvements.

b. M1A2 EUTE Background

In December 1991, TEC conducted ten force-on-force battles at Fort Hunter Liggett, California. The purpose of the battles (trials) was to operationally test the new M1A2 main battle tank and compare it to the current M1A1 main battle tank in different combat scenarios using actual soldiers. To accomplish this effort four scenario types were chosen: 1) Movement to Contact, 2) Deliberate Defense, 3) Hasty Defense, and 4) Hasty Attack. One trial of each scenario was conducted except for the Movement to Contact scenario, which had two iterations in different areas of the maneuver area. To ensure that results could be directly compared, each M1A2 trial had a corresponding M1A1 trial. Table I displays the scenarios and the number of combat systems on each side during the field battle [Ref. 16:p. 2].

Table I M1A1/M1A2 EUTE SCENARIOS AND FORCE SIZES

Scenario	#Blue Tanks	#Red Tanks	#Red I
Deliberate Defense	4	4	7
Hasty Defense	4	4	7
Hasty Attack	4	1	1
Movement to Contact I	4	2	3
Movement to Contact II	4	2	3

*Infantry Fighting Vehicle

c. Preparation of the Janus(A) Trials

In trying to replicate the parameters of the actual field test, TRAC-Monterey personally observed two of the test trials. By doing so, they felt confident that they had accurately portrayed the scenarios. In portraying the trials, several manual conversion steps were used, as well as automated conversion procedures.

(1) Conversion of Position Location Data

TEC provided Position Location (PL) files for every second of the respective vehicle's location during the battle. TRAC-Monterey converted the PL files to Janus(A) operational files, which included weapon systems for each side and their routes, by means of two Fortran programs.

(2) Vehicle Movement

The vehicle movements when simulated in Janus(A) did not generally match vehicle movement in the field trials. Specifically, while the automated conversion code traced the exact routes followed by specific vehicle, the synchronization of the timing of those vehicle over those routes in Janus(A) was often not similar to that of the actual field test timing. This occurrence is due to Janus(A) using an algorithm that relates individual vehicle speed to the maximum capability of the vehicle for the terrain features, specifically slope and vegetation, that the vehicle is traversing. Movement in Janus(A) is unable to consider the

input of the actual driver of the vehicles in the field test. Also, terrain resolution in Janus(A) may not allow for fine enough replication of the actual terrain variability. However, research currently being carried out at TRAC-Monterey, in conjunction with the Naval Postgraduate School, is intended to correct this deficiency.

There is an additional consideration for movement. In a field test, once a tank is killed its movement is terminated (an audio and visual cue is given to the tank crew notifying them that they are killed and they stop). The automated version of field movement to Janus(A) movement stops at that termination point. In a Janus(A) battle, that tank may not be killed at the same time and place as in the field trial. Vehicles that had terminated movement routes from the field trial had their routes extended in Janus(A) based on METT-T (Mission Enemy Task Terrain-Time) factors. The tank that was killed in the "open" in the field trial may be moved to a sheltered position by Janus(A) when the model is run. Therefore, the location of the tank in the field trial, when killed, may not correspond to its position in the model.

(3) Field of View

The fields of view of individual vehicles in stationary positions must be input to accurately reflect the battle. Janus(A) uses as a view the last direction the vehicle travelled. TRAC-Monterey used a combination of their

familiarity with the test conditions and METT-T factors to ensure that stationary vehicles had realistic views.

Janus(A) uses 50 meter cells (low resolution terrain data base) necessitating adjustments to vehicle emplacement and movement. This means that for every 50 square meters on the ground, Janus(A) considers the four corner points for elevation. This causes Janus(A) to be unable to capture the small undulations in the earth that can cover and conceal relatively large vehicles. The end result is that some vehicles that may be hidden from enemy view in the field test may be visible in the Janus(A) model. This problem was corrected by slightly moving some vehicles to take advantage of the vegetation in Janus(A).

d. Measures of Effectiveness (MOE) Selection

TRAC-Monterey considered detection ranges, engagement ranges, and system survivability as possible measures of performance. These measures were derived from the Critical Operational Issues and Concerns of the EUTE (Early User Test and Evaluation).

(1) Detection Range

Detection range was considered because of one of the M1A2's most significant fightability modifications, the Commander's Independent Thermal Viewer (CITV). The CITV enables the tank commander and gunner to look independently at the battlefield and search for targets. However, TRAC-

Monterey chose not to use detection range as the measure of performance because field test detections and Janus(A) detections do not apparently represent the same phenomenon.

Detection ranges were recorded in the test by means of a reconstruction of the gunner/commander communication. The time, target location, and tank location were all reconstructed from tapes of the actual test. The detection was credited only when a verbal or visual cue was given by the gunner/commander.

Janus(A) detections differ in that there is no crew involvement in the detection process: a detection is recorded when the weapon system's sensor detects and fully identifies the target. Therefore, TRAC expected that Janus(A) detection ranges would be generally longer than the field test detections.

(2) System Survivability

System survivability records the number of friendly survivors in each battle. TRAC did not choose system survivability because of the limited force size.

(3) First Engagement Range

First engagement range was considered because it has the potential to demonstrate the significance of the CITV to the M1A2. TRAC thought that using engagement range as an MOE would capture the effect of the independent views. TRAC also thought that they could capture the detection range

along with the views, but without the problems inherent in comparing Janus(A) and field test detections.

TRAC further limited their MOE to first engagement range. Theoretically, a trial with four blue tanks versus four red tanks could have a maximum of sixteen blue first-round engagements. TRAC imposed this limit to their MOE because subsequent engagements at the same target may depend on the first engagement. A basic assumption of each of the data analysis methods is that the data are independent.

e. EUTE Data Considerations

TEC's Range Measurement System (RMS), while gathering timely and thorough field test data, is not a perfect system. The PL data, while giving eight-digit PL to within 10 meters, has some inherent error. A realistic estimate is that actual vehicle positions are within a 30-meter radius of the PL data point.

Another possible source of error was due to TEC's engagement files not being complete because some engagements were not fully captured. An example is that some files report no target identification or range, but identify the firer. TRAC believes this may have occurred due to improper boresighting of the laser, RMS equipment malfunctions, or a missed shot. Approximately 21% of engagements were lost in this manner.

TRAC's concern was that the lost data may bias the results. If, for instance, the probability of an incomplete engagement was greater for engagements at greater range, the resulting data would indicate shorter mean engagement ranges than actually transpired. TRAC's belief in this regard was that lost engagements do not indicate bias, and were generally attributed to one tank during the battle that apparently did not correctly boresight its laser and could not record engagements at all ranges.

f. Experimental Design

TRAC's experimental design was intended to compare mean first engagement ranges of the field test and Janus(A) within each particular scenario and to determine if there were any statistically significant differences. TRAC could then "pool" their data, by scenario type, light conditions, and a cumulative Janus(A) versus cumulative field test. Each scenario was run three times.

The type of two-sample test that TRAC could do depended on the assumptions they could reasonably make. TRAC decided to first check the normality of the samples by using the Kolmogorov-Smirnov (K-S) Goodness of Fit test with a significance level of 0.05. This test indicated that only two samples, of the sixteen samples of Janus(A) and field trials, rejected the null hypothesis that the sample followed an approximately normal distribution. None of the four pooled

samples rejected the null hypothesis. However, due to the small sample sizes, TRAC did not have high confidence in the K-S test results.

TRAC decided to use the t test as one of their tests on the difference in the sample means because the majority of their data appeared to follow an approximately normal distribution. The t test relies on the underlying assumptions that the two sample populations are normal and that the two population variances are equal. However, comparison of sample variances showed significant differences in many cases. Despite TRAC's doubts in the validity of the assumptions of normality and equal variance, they decided to use the t test in their analysis and attempted to substantiate the results using non-parametric methods.

The non-parametric method chosen was the Mann-Whitney test. This method relies on the assumptions that both samples are random samples from their respective populations, and that in addition to independence within each sample, there is mutual independence between the two samples. For this experiment all of the assumptions were met. The Mann-Whitney test essentially combines the two samples being compared and ranks the values ordinally. The test statistic is then calculated using the assigned ranks, rather than the actual data values. This process effectively eliminates the effects of any underlying distributions associated with the data values.

3. Data Analysis

a. Results

Table II displays a comparison between the eight field trials and the sample of Janus (A) executions [Ref. 16:p. 7]. In all but two trials, the differences in the means of the first round engagements are not statistically significant.

Table II COMPARISON OF JANUS/EUTE TRIALS

Type	Scenario	Source	Sample Size	Mean Engage Range	Sample Variance	t test	M-W
MIA2	Movement to Contact I	Janus	12	1.57	.147	.15	.63
		Field	5	1.89	.185		
	Movement to Contact II	Janus	14	.39	.144	.68	.63
		Field	4	.36	.002		
	Deliberate Defense	Janus	27	2.04	.604	.001	.003
		Field	16	1.21	.353		
	Hasty Defense	Janus	9	1.19	.160	.157	.107
		Field	7	1.25	.918		
MIA1	Movement to Contact I	Janus	16	1.53	.063	.33	.057
		Field	7	1.29	.051		
	Movement to Contact II	Janus	16	1.42	.111	.21	.602
		Field	8	1.89	1.988		
	Deliberate Defense	Janus	24	.82	.144	.001	.002
		Field	12	1.45	.451		
	Hasty Defense	Janus	35	1.14	.177	.314	.868
		Field	6	1.32	.033		

b. Analysis of Deliberate Defense Differences

The trials whose mean first engagement ranges were

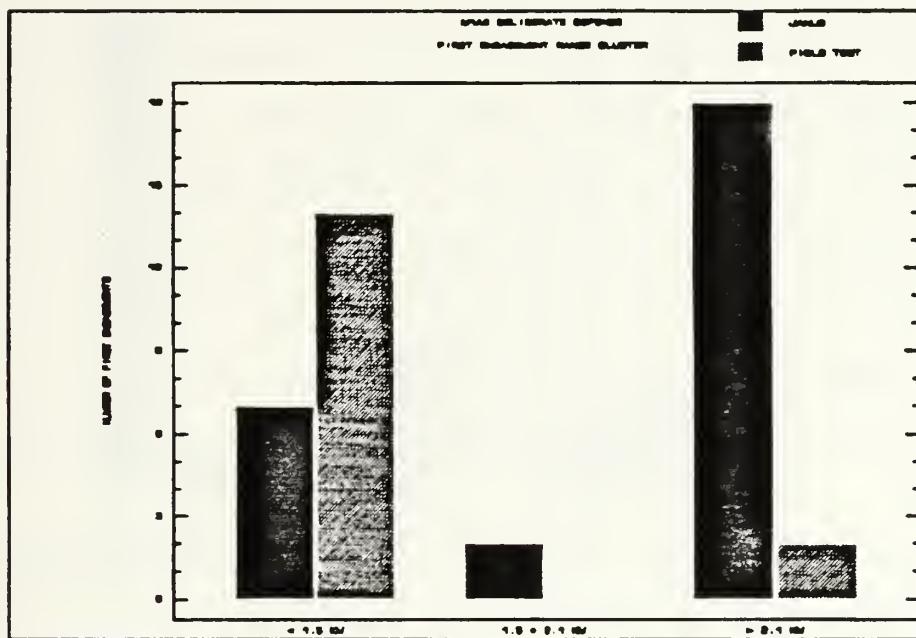


Figure 1 M1A2 Deliberate Defense Bar Chart

statistically significantly different between Janus (A) and the field test were the deliberate defense scenarios for both M1A2 and M1A1. TRAC surmised that there are three common factors leading to the differences in both trials. They are the terrain data base, line of sight from different defensive positions, and tactics.

In the M1A2 deliberate defense trial, the difference may be explained by an anomaly in the Janus terrain data base. TRAC analysis of the resulting data (Figure 1) shows that the blue forces in the Janus (A) engaged at a much greater range than they did in the field test [Ref. 16:p. 8]. The figure also shows that there were no field engagements

from 1500 meters to 2100 meters, while Janus(A) recorded many first engagements in this range band.

TRAC personally observed this trial at Fort Hunter Liggett and therefore had first-hand knowledge of the events. Soon after the battle began, the blue forces had line-of-sight to the red force attackers and fired several shots. Yet, the blue forces lost line-of-sight for a 600 meter period due to terrain masking, as the red force tactically used cover and concealment to close unobserved on the objective. This terrain masking occurred despite the blue forces having ample time to select the best defensive positions available in that area.

The abrupt changes in the terrain elevations that caused this occurrence are not captured in the 50-meter terrain data base by Janus(A). The blue forces in the Janus(A) M1A2 deliberate defense scenario had longer line-of-sight, particularly two blue tanks which took the vast majority of the longer shots. This occurred despite placing the tanks in the same grid location as they were in the field test.

The other trial in which a statistically significant difference occurs is in the M1A1 deliberate defense scenario. In this trial, the blue forces engaged targets at longer ranges in the field than in Janus(A)

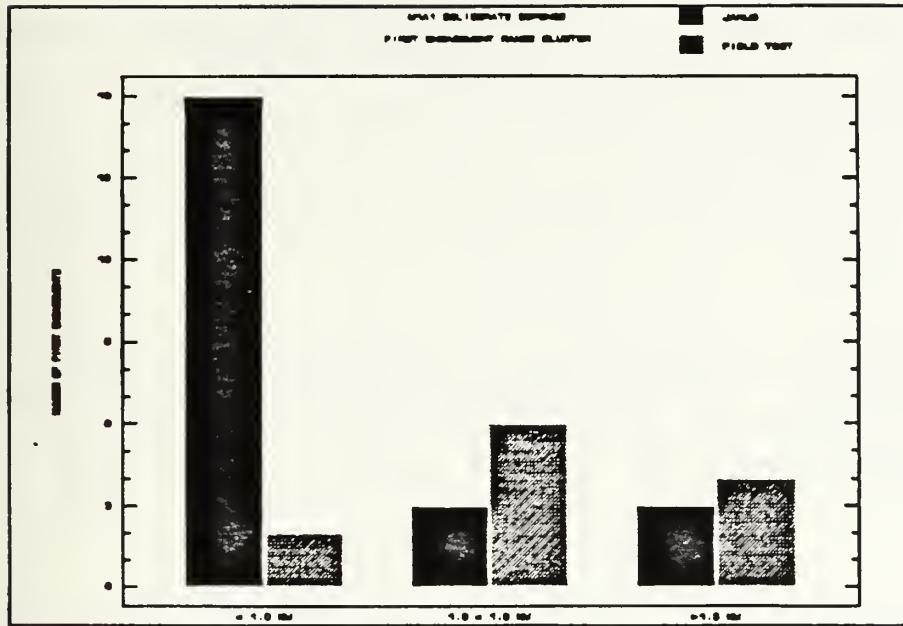


Figure 2 M1A1 Deliberate Defense Bar Chart

(Figure 2) [Ref. 16:p. 9]. The blue forces in the Janus (A) version of the M1A1 trial had minimal line-of-sight. Though TRAC-Monterey did not personally observe this trial, they were confident that in the field test, the blue forces chose defensive positions that allowed for optimal line-of-sight. The defensive positions in the Janus (A) M1A1 deliberate defense trial could not be improved with any minor repositioning. TRAC believes that the low resolution terrain currently employed in Janus (A) does not portray berms or sharp folds in the ground that would make good defensive positions.

c. Pooled Data Analysis

TRAC pooled certain scenarios that had an important common condition. TRAC pooled all movement to contact trials, all day trials, and all Janus(A) versus all field trials. There were no significant statistical differences in any of these comparisons. Table III summarizes the pooled movement

Table III POOLED MOVEMENT TO CONTACT FIRST ENGAGEMENT RANGE DATA

Scenario	Source	Sample Size	Mean Engage Range	Sample Variance	t test	M-W value
A1 MTC TRIALS	Janus	32	1.48	.087	.502	.515
	Field	15	1.61	1.1		
A2 MTC TRIALS	Janus	26	.93	.436	.328	.485
	Field	9	1.21	.741		
POOLED MTC TRIALS	Janus	58	1.23	.312	.191	.709
	Field	24	1.46	.973		

to contact trials, while Figure 3 graphically displays the similarity of pooled Janus(A) first engagements versus pooled field test first engagements [Ref. 16:p. 10].

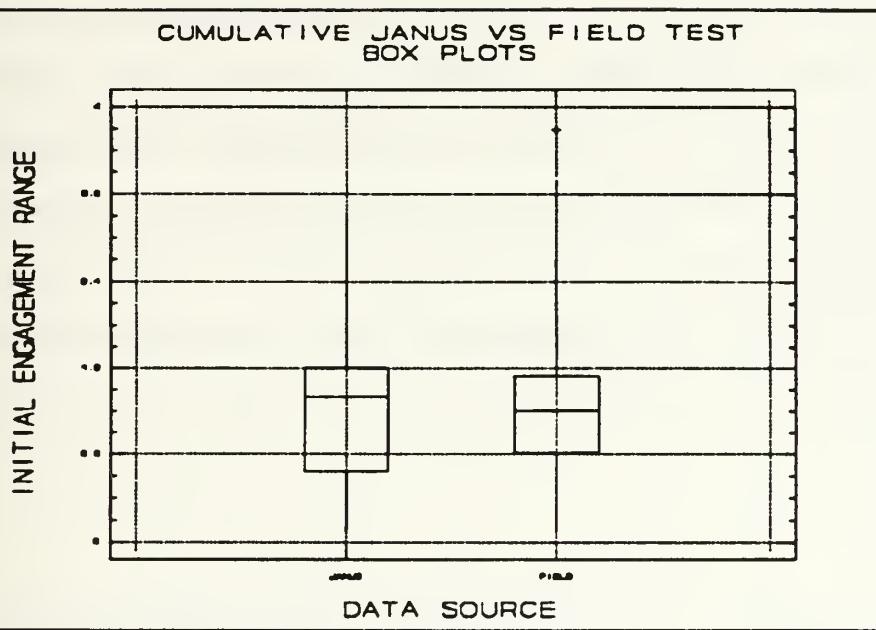


Figure 3 Box Plots of Cumulative Janus vs Cumulative Field Test

d. Conclusions

The majority of comparisons between Janus (A) and the field test trials is favorable. This indicates that Janus (A) may acceptably portray the M1A2 and M1A1 tanks. TRAC cautiously recommended that Janus (A) be accredited for Model-Test-Model for M1A2 and M1A1 first engagements. TRAC's reservations remain due to the terrain database limitations of Janus (A) and the small sample size of the field test data.

TRAC believes that there are solutions to both the Janus (A) terrain database limitations and the small sample size. The integration of a one-meter terrain database,

currently available for Fort Hunter Liggett, into Janus(A) should make more accurate line-of-sight calculations, and subsequently more realistic engagements. This is a research project directed by TRAC-Monterey in Fiscal Year (FY) 93. Additionally, future field tests will benefit from pre-test modeling, which will optimize the test design. Pre-test modeling can focus the test trials in certain scenario types or conditions that will more readily show differences between the new and old weapon system and will provide a larger sample size.

D. SUMMARY

Model-Test-Model is a concept to assist the operations research analyst and modeler in verifying and validating computerized combat simulation models. It also provides data to the operational tester that is accurate enough to be incorporated in the testing process. The successful application of a M-T-M to testing produces two desirable products: first, a comprehensive evaluation of a system with credible data, and second, a reasonably well-calibrated model of a system that can be used for extrapolations and "what if" investigations. The overall goal of M-T-M is to save the Army time and money by using a validated/accredited model to provide operational test data. This operational data is used in conjunction with field test data to provide a better

assessment of the test and to make more accurate decisions earlier on in the acquisition process.

The pretest phase of M-T-M will benefit the tester by providing insight on test design. The proponent agency may also profit by analyzing pretest results and modifying tactics if necessary. The post-test phase of M-T-M allows the modeler to benefit from the accreditation process and the evaluator is able to then use the accredited model to extend test results.

V. THE FUTURE OF TEST AND EVALUATION (T&E)

A. INTRODUCTION

This chapter looks at the future of operational testing as we approach the 21st century. This chapter also provides an analysis of the changes that will need to be made in the T&E community as a result of a new defense environment. Finally, some recommendations are made of how the testing community should react to these changes. But before these issues are analyzed, further background into why the changes are being brought about is necessary.

B. CHANGES IN THE DEFENSE ENVIRONMENT

A lesson learned from Desert Storm is that our prior focus on strategic nuclear forces and conflicts in central Europe did not adequately prepare us for likely future regional conflicts. In these future conflicts we will be facing state-of-the-art equipment from many nations-including our own. It was clear that we will need the capability for locating and destroying mobile targets (e.g., SCUD missile launchers), for all-weather precision-guided weapons, and for real-time intelligence data available on front line ships, planes and tanks [Ref. 6:p. 10]. The information technology explosion of the end of the 20th century will spread rapidly throughout all

nations of the world over the next few decades, providing even small, Third-World forces with incredibly sophisticated, destructive capabilities.

Historically, The United States has counted upon "technological superiority" to achieve military leadership. The future challenge for the nation will be to maintain technological superiority in its deployed forces. This must be accomplished while surviving on a much smaller defense weapons budget.

Maintaining technological superiority with a smaller budget presents a very difficult challenge. If we look back for the past 45 years, we see that, from generation to generation of weapon systems, costs have been rising to match the increasing weapons performance [Ref. 6:p. 10]. In the future we are going to be demanding increasing performance, but at a far lower price. In addition, we will be buying fewer weapons, driving up per-unit production costs. In order to realize technological superiority in our deployed forces, advanced technology must move rapidly through the acquisition process. However, our past performance in weapons acquisition shows that a system takes, on average, over 16 years to field. This situation is inconsistent with the rapid development of new technologies.

Weapons costs and time trends are further exacerbated by mounting problems in the U.S. defense industrial base. The most visible sign is the hundreds of thousands of people being

laid off in the defense industry. Rather than some rational down-sizing and restructuring strategy for achieving future efficiency and effectiveness in the industrial base with the smaller budget, firms and unions have been lobbying their congressional representatives for more "pork and protectionism." Congress has responded by adding billions of dollars for procurements of old weapon systems--many of which were not even requested by DOD. The approach seems to be an attempt to maintain an increasingly inefficient and ineffective industrial base. This will drive weapons costs still higher and will stretch out programs still further. A change is needed!

Dr. Gansler (former deputy assistant secretary of defense and the senior vice president of the Analytic Sciences Corporation) suggests that there are four areas that represent the largest potential for realizing far greater national security with far fewer dollars. He warns that all four areas are "countercultural" and therefore will be extremely difficult to implement and will require strong, sustained leadership. Specifically, these four new directions call for the:

- Dramatic restructuring of the weapons requirements process.
- Similar dramatic restructuring of the defense DT&E arena.
- Total restructuring of the defense industrial base.

- Dramatically streamlined, and yet upgraded, government and industry acquisition corps.

A brief discussion will be made of each of these areas. Then their impact on the T&E community will be analyzed.

1. Weapons Requirements Process

Of primary importance is the military requirements process. This is what drives the acquisition process. However, weapons requirements have been historically driven by a desire to design the highest-performing weapons systems in the world. Later would come the thought process of how to build these weapons and, if possible, how to slightly reduce the cost of very expensive weapons. If we contrast this military requirements process with the commercial world, we find that many new, high technology systems are also striving for state-of-the-art performance. However, the commercial world also has a firm requirement for low production and support costs. This is the difference that the defense world must remove. In the future, the DOD requirements process must be driven by the simultaneous need to improve performance and lower production and support costs. To achieve lower costs in our weapons means that we must frequently look to nontraditional approaches. The traditional approach to weapons acquisition, i.e., writing a requirement for a new weapon and beginning the development of a full, new system, is more likely to be replaced with the far less costly approach

of upgrading existing platforms and weapons with more advanced information technology.

2. Major Changes in the Research and Development (R&D) Process

Changing/improving the requirements process necessitates changes in the R&D process. More focus will be on the upgrading of existing weapons and platforms with improved information technology subsystems. There will also be other major changes in the RDT&E (Research, Development, Test and Evaluation) process.

The primary change will be increased allocation of the limited defense resources to the "front end" of the acquisition process. Since production of current weapons is so expensive, the choice for the future appears to be to invest in R&D. This will allow the nation to keep its technological edge. Then a decision will have to be made to place a few of these R&D systems into production, but only when they represent a quantum leap forward in operational military effectiveness. We can expect to see far more use of weapon and subsystem prototypes in the future--with the explicit expectation that many of these will not go into full development and production. Rather, instead of prototypes being used solely for technical feasibility demonstrations, they will also be used for demonstrating affordability. This

results in a very different definition of "prototypes" than has been used in the past.

3. Combined Industrial Base

In the area of defense industrial base, Dr. Gansler fully expects the nation will no longer be able to afford, nor desire to have, a unique defense industry. Rather, he expects to see a largely integrated civil/military industrial base. Dr. Gansler writes that three recent shifts in technology have brought about this possibility: 1) many commercial parts and technologies now exceed military requirements and have higher performance and far lower costs than their military counterparts; 2) modern "flexible" manufacturing technology allows for different military and commercial products to be built on the same line, as long as the production processes are similar; and 3) critical defense technologies are increasingly overlapping with technologies that are essential for future commercial competitiveness. Thus, Dr. Gansler envisions that future products will be designed to meet dual-use requirements and will be designed to be built in dual-use factories, using largely commercial parts, materials, and software. This will require dramatic changes in such areas as cost accounting and auditing, military specifications and standards, and procurement practices. Such changes are necessary for the DOD to achieve the efficiencies and state-

of-the-art technologies needed for an affordable, yet effective, military.

Dr. Gansler's views are consistent with those stated in a January 1992 DOD Briefing by the Deputy Secretary of Defense. The meeting was attended by the Secretary of Defense and the Chairman of the Joint Chiefs of Staff. His views are also consistent with those espoused by Former Under Secretary of Defense Don Yockey in his May 1992 memorandum on Defense Acquisition. His memorandum was distributed DOD-wide to include the secretaries of the military departments and the Chairman, Joint Chiefs of Staff.

4. Professional Acquisition Corps

The defense world will have to move toward streamlined organizations and more highly trained and skilled personnel. These skilled people must be able to have management decisions delegated to them and can be empowered to take on the added responsibility associated with that decision-making. This means that, on both the government and industry sides, the key people in the acquisition community of the 21st century are going to have to be highly experienced and educated. There will be far fewer of these key people. Yet, they will be given the tools (such as advanced computing, communications, and display capability) that will allow them to do their work far more efficiently and effectively.

The Army has made great strides in this area by forming the Army acquisition Corps. The acquisition corps will consist of dedicated, professional acquisition personnel. These personnel will serve throughout their careers in the acquisition community. They will also receive intensive acquisition education and training. This sharply contrasts with the old method of putting nondedicated people in Program Management positions and letting them "sink or swim". This method historically led to weapon system programs being behind schedule and over budget. The acquisition corps is an attempt to mitigate this problem.

C. ANALYSIS OF IMPACT OF CHANGES TO T&E

How will the changes in the defense environment affect T&E? One of the most dramatic impacts comes from lessons learned in the commercial world as a result of increasing international competition. Japanese firms have shown that the most efficient and effective method of developing and deploying the most advanced products is to use a "continuous product and process improvement" approach. This means that you don't plan on putting every new technology into the system right from the start. Instead, as each technology is proven, you continuously modify the products, and you, simultaneously, continuously improve the design, production, and support processes to improve the reliability and lower the costs of the product. For the defense community, this changes the

whole acquisition cycle concept. For the T&E world, it causes a dramatic change in outlook. It actually returns T&E to its original function of being a part of a continuous development process, rather than an "independent auditor" of the development process. The T&E focus must be on evaluation (versus "testing to see if it is acceptable") since the T&E community must always be looking for ways to improve performance at lower costs. It becomes important to have failures in the test process. This is the only way the testing community can learn, and continue to improve. The goal is to test systems beyond their limits, in order to see what can be done to make them even better in the next round.

As with the other RDT&E changes, the new emphasis on prototypes that are being evaluated for the combination of technical feasibility, affordability, and operational utility, places a far greater burden on the testing community. To reduce cost systems, the testing community must assess which performance parameters are the cost drivers, which are essential to significantly enhance operational utility, and which are just nice to have. In essence, the testing community becomes an essential element in the affordability trade-offs between cost drivers and performance drivers. The testing community must also judge whether or not the prototype results are "scalable" to the operational production versions of the system, and to the likely operational scale threats that may be encountered. Finally, the T&E community will be

tasked to help make the most critical decision in terms of the new acquisition cycle. The decision is to determine whether or not these prototypes, if developed and produced, would represent a substantial improvement in military force effectiveness.

In the past, the most critical decision for the testing community did not occur until after the weapon system had gone through full-scale development and was essentially ready for production. This new, earlier decision will be a more difficult one. However, part of the rationale for these new prototype developments is to maintain a core engineering and manufacturing defense industrial base for next-generation weapon systems. If the T&E community decides that a particular prototype does not represent a substantial improvement, the requirement will be sent back to the R&D community for another cycle. The R&D community will then use the next generation of technology which will, hopefully, provide the needed substantial enhancements in either lower cost or higher performance.

Three developments occur from this new weapons acquisition concept. First, a much more rapid cycle time from the initiation of a concept to the field deployment of a prototype is needed to determine if a concept is worthy of subsequent development and production. Second, the testing community must be able to take the design from prototype to production rapidly--if we decide to produce it. This will enable the

nation to maintain technological superiority of fielded systems. Third, the testing community will most likely have a much more highly integrated development test and operational test program on these prototypes since there will be simultaneous testing for technical feasibility and operational utility.

Along with the new acquisition process placing a greater burden upon the testing community, so will the new industrial base strategy. The use of far more commercial parts, subsystems, and software means that more testing of these elements will be required to ensure that they meet the military's need. They will now have been designed and built to commercial specifications and standards, not traditional military specifications and standards.

1. Information Technology

The impact of supercomputing, advanced displays, and next-generation communication technology will be seen first in the weapon systems and the command, control, communications, and intelligence arenas. This will have the effect of dramatically changing not only weapon systems themselves, but even how warfare will be fought in the 21st century [Ref. 6:p. 13]. For the testing community, this will cause a significant shift from the current focus on weapons testing to a far greater emphasis on "operations testing". Operations testing will tightly link the command, control, communications, and

intelligence arena into the weapons operations themselves. It will also cause a significant shift from traditional hardware testing to software testing (a much more difficult task).

2. Simulation

The impact of the information technology, coupled with a shrinking budget, is the move toward more reliance on simulation and modeling (Chapter III gives a detailed account of the Model-Test-Model concept which will be an integral part of the modeling effort for T&E). One can envision the shift in the T&E regime to that of live testing a system only to validate the models. While there will be heavy reliance upon simulation and modeling, it is clear that it will be absolutely essential, in all cases, to run a limited number of tests to achieve total confidence in the validity of the models. It will be essential for the simulations and models themselves to be "validated" and written to be reusable and transportable. They will be used throughout the weapon system's life cycle. They will be used for establishing the requirements, then in the preliminary design phase, subsequently, as the system evolves, and eventually as it is upgraded. Throughout this process, as tests are run the models will become more valid. They will rely on complementary weapons systems, rather than on the system under development for this further validation. We can see that modeling and simulation are not alternatives to weapons and

subsystem testing, but rather necessary complements, which will improve the effectiveness and efficiency associated with testing.

Simulation and modeling will also be used to improve the T&E process. Through the use of CIM (the Corporate Information Management system that is being developed to improve data processing within the government), CALS (the Computer-Aided Acquisition and Logistic System that will directly couple the government and industry information systems), and EIP (the Enterprise Integration Program that will link together the information systems within firms and between firms in the industrial structure), it should be possible to dramatically reduce the time and level of manpower associated with the overall T&E process [Ref. 6:p. 14]. As numerous recent studies have demonstrated, T&E is a (if not the) major cause of the excessively long acquisition cycle for weapons today [Ref. 6:p. 14]. To be fair, these studies have also shown that the majority of the time is not taken up with testing, but in "waiting," preparing documentation, and other low-value activities that can be dramatically reduced through the application of advanced information technology. In the future, the program office, the testing community, and the contractors involved will all be "on the net" whenever required and will be able to reach agreements collectively in days instead of weeks or months (the testing community has

made a major stride in this area with the establishment of TECNET (Test and Evaluation Community Network)).

3. Future Changes Require New Tools

The testing community can look forward to assistance in the new computer-based tools being developed. These tools include "expert systems" to assist in evaluating test results. These tools also include the Defense Advance Research Projects Agency (DARPA)-sponsored "Case-Based Reasoning" models, which will aid a program office in researching relevant historic test programs. The developed support tools and simulations and models will have to be as generic as possible, but their funding will require significant process (R&D) dollars. If the funding for these tools is not realized then the long-term benefits that they provide will be lost.

4. Increased Diversity of Systems to be Tested

The testing community can expect to see a far greater variety of systems to test. This is due to the manufacturing industry shifting toward flexible manufacturing systems, combined with a move toward a continuous product and process improvement cycle. There may be only a few of each type of system, and because of their user friendliness this will be no significant problem for the operators or maintenance people. However, such variety will tax the testing community.

5. Impact on T&E Personnel

The changing defense environment and its impact on the testing community will necessitate the need for extremely high-quality government and industry T&E personnel. The transformation taking place will require that the testing community no longer focus on "go/no-go" decisions, but rather that they apply management judgement in evaluating the systems under test [Ref. 6:p. 14]. The concept of "technically acceptable, low bid wins" is incompatible with "world-class" operations, and the testing arena will be no exception. Therefore, to be a "world-class" operation the testing community has to recruit, train, and contract for only the best.

D. HOW THE TEST AND EVALUATION COMMUNITY SHOULD REACT TO CHANGES IN THE DEFENSE ENVIRONMENT

To handle the changes being brought about in the defense environment, the T&E community must understand its function. The T&E function is to provide information on which decision-makers can rely with full confidence in making acquisition decisions [Ref. 15:p. 13]. To do this, the testing community must conduct sound tests, evaluate the results carefully and provide reports that are clearly written and free from bias. Organizational structures must not be allowed to coerce inadequate testing or biased reporting; but at the same time, the testing community cannot so distance itself from the

developmental process that they do not understand the system under development. Additionally, the testing community must be able to competently evaluate the utility for T&E purposes of data that is generated during the development process. All of this must remain true after the testing community accommodates the changed environment in which T&E must take place.

The testing community must respond to its changing environment in two ways [Ref 15:p. 13]. The first entails doing the routine T&E functions efficiently and effectively. The second involves finding new and better ways of operating. These two methods are discussed in further detail.

1. Routine Functions

Among the functions to be done more effectively, the testing community must:

- a. Make sure that lowered budgets do not mean inadequate T&E.**

The testing community needs to be very careful not to allow its standards to be compromised. Program managers, and possible even decision-makers, will want to have effectiveness issues resolved with lower testing. The testing community must resist the pressure to eliminate testing that it believes to be important. If not supported in these matters, the testing community must be forthright in

establishing a written record that makes clear who made what decision.

b. Integrate Life-Cycle Management.

In the past when program executive officers and program managers got less funding than they knew they needed, the elements of the program that were "axed" first tended to be training, documentation, and logistic support. Until reinstitution of life-cycle management or "cradle-to-grave" program oversight by the same organization, this situation will continue. This is because the "buyer" knows he/she will not be accountable later for in-service support measures.

c. Be watchful of the user's interests in coordinating on Test and Evaluation Master Plans (TEMPs).

With lessened participation by the user in the acquisition process, the testing community must ensure that an adequate TEMP is produced. It must ensure that the TEMP and derivative test plans cannot lead to satisfactory evaluations for systems that do not meet the needs of the soldiers. When staffs of senior officials "lean" on the testing community to do something that is believed to be incorrect, ensure that it is made a matter of record so that there is an audit trail.

d. In Development, Test and Evaluation (DT&E) be cognizant of the technological and engineering capability of manufacturing to produce the system.

This is so that the testing community is not caught unaware of the decision to begin production. The testing community has been not been vigilant in this area in the past. The possibility of short-fused orders to begin production means more and better scrutiny in the future.

e. Get involved in every project as early as possible.

The earlier the testing community is involved, the better for the whole program. This has the effect of being able to alert the program manager of potential T&E problems early in the program. It also allows the testing community to gather more data on the system in order to have a better and shorter T&E period.

2. New Functions

To survive in the new defense environment, the testing community needs to conduct T&E competently. The testing community also needs to show the acquisition officials that it (testing community) is ready and able to do so in ways that serve the officials' needs in the new defense environment. The new environment demands that the testing community be able to respond to an opportunity to conduct an operational assessment, DT&E, or, if the system is sufficiently mature,

OT&E (Operational Test and Evaluation), in field circumstances, on short notice. The Secretary of Defense must be confident that the testing community can respond promptly and effectively so he will want to do the T&E that should be conducted before making the decision to produce a design prototyped earlier. This can be achieved by a change in the T&E cultural mindset. Specifically, the testing community must not be thought of as the final exam at the end of the process, but as an integral part of the process from the very beginning. To accomplish this, the testing community must:

- a. Enable the test project officer to be able to respond immediately to an opportunity to test.*

This can be done only by developing new internal procedures to facilitate such a response. The testing community has the talented people to do this if those people are empowered to do so. This includes supporting them with authority, resources, and test formulation software that could reduce the necessary lead time to 24 hours or less.

- b. Form pre-designated deployable T&E teams.*

The deployable teams would be organized and equipped to respond within a few hours to an opportunity to test. They should be provided with the ability to establish voice communications with the test project officer.

c. Preplan the movement of test teams and equipment.

This would involve looking at testing on a global scale. The testing community must be cognizant of opportunities to test anywhere a system may be deployed.

d. Establish and sign Memoranda of Agreement (MOA) with commanders of theaters where testing may occur.

The MOAs should define the responsibilities of the person(s) designated by the theater commander to conduct the tests. The responsibilities of the T&E teams must be provided by the T&E organizations. The MOAs must also establish the authority of the T&E teams to operate independently of coercive influences. It must define measures to protect the security of the test data and the procedures to be followed in emergencies.

e. Establish field data protocol.

Means of data reduction and analysis between the field teams and the test facility must be established. This would enable data analysis to begin as soon as testing begins. Procedures need to be enacted to provide for the completion of evaluation, drafting the report, and review by the test team before the report is released, regardless of the location of the test team.

f. Demonstrate the T&E changes to the acquisition officials.

Acquisition decision-makers must be made aware of the commitment of the testing community to be an efficient and competent part of the acquisition cycle. Action must also be taken to update applicable DOD and service directives to reflect the testing community's new capabilities.

E. SUMMARY

The operative word in the defense community is *change*. The Department of Defense can no longer afford to do "business as usual" and expect to survive. In response to decreased personnel and budgets, defense organizations will either adapt and change or die. This is particularly true of the T&E community. The acquisition process must change and is changing to meet the realities of today and tomorrow. The T&E community cannot afford to take a "wait and see attitude", but must take a proactive approach of dealing with the coming changes. This means that instead of being the "final exam" at the end of the acquisition cycle, the testers have to be an early and on-going part of the process.

VI. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

The underlying theme in the acquisition world is *change*. This is reflected in the trends in the defense industry. These trends include a changing threat, reduced budget, consolidation of resources, fewer systems, etc. To quote Bob Costello, the keynote speaker at the International Test and Evaluation Association (ITEA)/American Defense Preparedness Association workshop in Las Cruces in March 93, "Change or die" [Ref. 2:p. 3]. He predicted adamantly that if an organization isn't in Washington understanding the changes and hurrying back home to make the necessary adjustments in his or her organization, that organization may not be around very long. Although some might think Costello overstated his point, few would disagree that there is far more pressure to change.

This statement is the cornerstone for the need for change in Test and Evaluation (T&E). There is a strong need for more T&E on the front end of acquisition. The focus of T&E must shift from one of testing to see if the system is acceptable to one of evaluation. This is necessary to ensure that a system, if developed and produced, will provide a substantial increase in military force effectiveness. The Defense

Department has shown an interest in early assessments and the desire to start planning for the evaluation as early as possible. This is the change the testing community must plan for--the necessary and critical emphasis on evaluation. No one is suggesting that the T&E community give up testing, but everyone involved in testing needs to advocate that a much better job be done of evaluation.

A number of things are happening that necessitate this reemphasis on evaluation. With the reduced budget, the testers will have to demonstrate, before funds are committed, that a new system can make a contribution. This is the reason for the interest in "distributed interactive simulations" where new concepts can be evaluated long before metal is bent. We will also have less funding for new test capabilities while at the same time we are faced with more complex technologies to test. All this requires more up-front thinking and planning.

The testing community needs to get serious about minimizing duplication between development and operational testing, using modeling and simulation, and taking advantage of other areas where we can reduce testing cost and increase evaluation efficiency. The testers need to define an evaluation framework early in a program--i.e., take a systems's approach to T&E (Dr. Deming's approach to quality). The testing community has long recognized that such a framework is needed, and has discussed its importance in

forums such as the combined ITEA and Military Operations Research Society workshop on "Emphasizing the 'E' in T&E. However, there are few examples of an integrated, comprehensive evaluation framework using a combination of contractor, development, and operational testing to optimize a T&E program from a total system's perspective. The testing community has talked about this for a long time. The T&E community must be proactive in changing the defense environment and in implementing the changes it knows it must make.

B. RECOMMENDED AREAS OF FUTURE RESEARCH

Areas recommended for future research include the possibility of consolidating some T&E activities in order to reduce operating costs. Another area for future research would entail developing a database that could be used to provide a detailed list of the resources available throughout the T&E community.

APPENDIX A. ABBREVIATIONS AND ACRONYMS

A	Ampere
ADF	Automatic Direction Finder
ADFS	ARMVAL Direct Fire Simulator
ADWFS	Advanced Development Weapons Fire Simulator
AFB	Air Force Base
AHIP	Army Helicopter Improvement Program
AM	Amplitude Modulation
AMC	Army Materiel Command
AMSAA	Army Material systems Analysis Activity
ARMVAL	Advanced Anti-Armor Validation
ASCII	American Standard Code for Information Interchange
BIT	Built-In-Test
CALS	Computer-Aided Acquisition and Logistics System
CAT	Computer-Aided-Test
CDEC	Combat Developments Experimentation Command
CDL	Computer Data Link
CIM	Corporate Information Management
CITV	Commander's Independent Thermal Viewer
CPX	Corps Post Exercise
DARPA	Defense Advanced Research Projects Agency
DEC	Digital Equipment Corporation
DES	Data Encryption standard

DFS	Direct Fire Simulator
DIVAD	Division Air Defense
DOD	Department of Defense
DRC	Data Reduction Center
DT	Development Test
DT&E	Development Test and Evaluation
EAB	Extended A-Station to Micro-B
EBA	Extended Micro-B to A-Station
EBCDIC	Extended Binary Code Decimal Interchange Code
EIA	Electrical Industry Association
EIP	Enterprise Integration Program
ELOSS	Engagement Line-of-Sight System
EMI	Electromagnetic Interference
EUTE	Early User Test and Evaluation
FAA	Federal Aviation Agency
FDTE	Force Development Test and Evaluation
FEBA	Forward Edge of Battle Area
FHL	Fort Hunter Liggett
FOM	Figure of Merit
FM	Frequency Modulation
FTX	Field Training Exercise
FY	Fiscal Year
GaAs	Gallium Arsenide
GAS	Gun Azimuth System
GHz	Gigahertz
GOES	Geostationary Orbiting Environment Satellite

Hz	Hertz
I ² C ²	Integrated Information Control Center
ICN	Instrumentation Computer Network
IDFSS	Infantry Direct Fire Simulator System
IFCAS	Indirect Fire Casualty Assessment System
IFV	Infantry Fighting Vehicle
ILS	International Laser System
I/O	Input/Output
IPS	Instrumentation Power System
IRIG-B	Instrumentation Group, B
ITEA	International Test and Evaluation Association
KHz	Kilohertz
KBPS	Kilobits-per-second
KM	Kilometer
K-TOPS	K-Band TRADOC Obscuration Pairing System
LCC	Life-Cycle Cost
LSIS	Laser Spotting Information System
LTA	Linear Triaxial Accelerometer
MAFIS	Mobile Army Field Instrumentation System
MAIS	Mobile Automated Instrumentation Suite
MBPS	Megabits-per-second
MCC	Main Computer Complex
MCU	Microcomputer Unit
METT-T	Mission Enemy Task Terrain-Time
MHz	Megahertz
MILES	Multiple Integrated Laser Equipment System

MM	Millimeter
MMCS	Mobile Multi-Purpose Control Station
MOA	Memorandum of Agreement
MOE	Measure of Effectiveness
MOPP	Mission Operational Protective Posture
MRAD	Milliradians
MS	Millisecond
M-T-M	Model-Test-Model
MUX	Multiplex
mV	Millivolt
NDI	Nondevelopment Item
NM	Nanometer
NSA	National Security Agency
OP-AMP	Operational Amplifier
OPTEC	Operational Test and Evaluation Command
OSD	Office of the Secretary of Defense
OT	Operational Test
OT&E	Operational Test and Evaluation
OTEA	Operational Test and Evaluation Agency
PL	Position Location
PLB	Programmable Logic Box
POSNAV	Position Navigation
RCS	Range Communication System
R&D	Research and Development
RDT&E	Research, Development, Test and Evaluation
RF	Radio Frequency

RIOTS	Range Measuring System Input/Output Test Set
RMS	Range Measurement System
RTCA	Real-Time Casualty Assessment
RTCC	Real-Time Computer Controller
RTS	Range Timing System
SCOM	Short Communication Message
SEO LGL	Schwartz Electro-Optical Large Gun Laser
SEO SGL	Schwartz Electro-Optical Small Gun Laser
SPIPS	Serial Programmable Instrumentation Pallet System
SPLB	Serial Programmable Logic Box
SS/LEI	Speech Synthesizer/Limited Expansion Box
SYNC	synchronization
TCG	Time Code Generator
TDY	Temporary Duty
T&E	Test and Evaluation
TEC	TEXCOM Evaluation Command
TECNET	Test and Evaluation Community Network
TEMP	Test and Evaluation Master Plan
TEXCOM	Test and Experimentation Command
TRADOC	Training and Doctrine Command
UCT	Universal Coordinated Time
UHF	Ultra-High Frequency
ULM	Upgraded Logic Module
UTS	Upgraded Transmission Signal
VDR/DC	Video Data Reduction/Debriefing Center

VGVA	Variable Gain Voice Amplification
VHF	Very High Frequency
VIDS	Visual Information Display system
VIMS	Video Instrumentation Mobile System
VTR	Video Tape Recorder

LIST OF REFERENCES

1. Bundy, Dennis D., Generic Model-Test-Model Using High Resolution Combat Models, TRADOC Analysis Command-Monterey, June 1991.
2. Costello, Bob, The Changing Defense Environment, The ITEA Journal of Test and Evaluation, June 1993.
3. Defense Systems Management College, Test and Evaluation Management Guide, March 1988.
4. Department of the Army, Operational Test and Evaluation Command Interim Policy Guidance 92-3, Reporting Tests, Evaluation, and Assessment, Alexandria, VA, 13 August 1992, (Draft).
5. Gansler, J.S., National Security and T&E at the Beginning of the 21st Century, The ITEA Journal of Test and Evaluation, March 1993.
6. Greenlee, Don, "Why Do We Test? (When Can We Stop?)", Test and Evaluation Symposium VII, 26 February 1991.
7. Hoeber, Francis P., Military Applications of Modeling: Selected Case Studies, Gordon and Breach Science Publishers, 1982.
8. "Improved Assessment Procedures Would Increase the Credibility of Results," GAO DOD Simulations, December 1987, p. 13.
9. Interview: Chief, Instrumentation Engineering Division, Fort Hunter Liggett, 17 June 1992.
10. Interview: Director, Operation Instrumentation and Testing, Fort Hunter Liggett, 17 June 1992.
11. Instrumentation Handbook 3rd Edition, TEXCOM Fort Hunter Liggett, CA, 1988.
12. Low, L. J., Theater-Level Gaming and Analysis Workshop for Military Force Planning (Concept and Plan), Stanford Research Institute, May 1977.
13. Marine Corps Operational Test and Evaluation Activity, Operational Test Project Officer Methodology Guidebook, July 1989.

14. Office of the Director of Operational Test and Evaluation, Operational Suitability Guide: Volume I-A Tutorial, February 1990.
15. Parker, Ted, A Time For Action: New Defense Environment Requires Changes in T&E, The ITEA Journal of Test and Evaluation, June 1993.
16. Paulo, Eugene P., Comparison of M1A1/M1A2 Battle Results Between Janus(A) and an Operational Field Test, TRAC-Monterey, January 1991.

INITIAL DISTRIBUTION LIST

	No. Copies
1. Defense Technical Information Center Cameron Station Alexandria VA 22304-6145	2
2. Library, Code 052 Naval Postgraduate School Monterey CA 93943-5002	2
3. Program Executive Officer Communication Systems ATTN: SFAE-CM Fort Monmouth NJ 07703-5000	2
4. U.S. Army OPTEC ATTN: CSTE-MP Park Center IV 4501 Ford Avenue Alexandria VA 22302-1458	1
5. TEXCOM Experimentation Center Director, Operation, Instrumentation, and Testing ATTN: CSTE-TEC-T Fort Hunter Liggett CA 93928	1
6. U.S. Army White Sands Missile Range ATTN: STEWS-MTD-BP (Mario M. Correa) White Sands Missile Range NM 88002-5157	1
7. Professor Nancy C. Roberts (Code AS/Rc) Naval Postgraduate School Monterey CA 93943-5100	1
8. Professor David R. Henderson (Code AS/Ht) Naval Postgraduate School Monterey CA 93943-5100	1
9. Dr. David V. Lamm (Code AS/Lt) Naval Postgraduate School Monterey CA 93943-5100	3
10. CPT Lawrence L. Turner, Jr. 1002 Creek Street Copperas Cove TX 76522	1

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE SCHOOL
MONTEREY CA 93943-5101



GAYLORD \$

DUDLEY KNOX LIBRARY



3 2768 00307455 0